THE UNIVERSITY
of EDINBURGH

Investigating the 11q23.1 Colorectal Cancer Risk Locus

Ruby Tamara Osborn

THE UNIVERSITY *of* EDINBURGH

PhD

The University of Edinburgh

2019

For Thom

Abstract

Colorectal cancer (CRC) is the fourth most common cancer in the UK and the second highest cause of cancer deaths. It is a complex disease with multiple environmental and genetic factors. Some risk genes have a low frequency of the risk allele in the population, but have a high effect, and have been well studied. Others are more common but have a much smaller effect; many of these are still being identified and characterised.

The 11q23.1 locus was linked to CRC risk by a genome-wide association study, via the single nucleotide polymorphism rs3802842. Subsequent expression analysis on human colonic tissue revealed the risk allele of rs3802842 is correlated with lower expression of three adjacent genes: *C11orf53*, *C11orf92* and *C11orf93*. *C11orf53* and *C11orf93* are protein coding, while *C11orf92* is non-coding. All three are of unknown function. This locus is not associated with risk of cancer in any other tissue, but has been linked to colitis-associated-CRC in mice.

To investigate these genes and their role in CRC, I have examined their localisation in intestinal tissue in humans and mice. I have created a knockout mouse model to study the effects of the loss of the genes on development and CRC induction, and used ex-vivo systems for further functional assays.

The three genes have distinct transcript patterns along the length of the intestines, which are consistent between humans and mice. These patterns do not indicate they are specific to either stem cells or differentiated cells, although all genes show higher expression in the base of the crypt than the top. There are several significant differences in transcript level and distribution in the intestinal tissue of an established mouse CRC-susceptible background ($Apc^{Min/+}$) compared to wild type tissue. Previous work by Claire Smillie and evidence that I have gathered using protein localisation indicate that at least one of the genes in this locus may play a role in the endoplasmic reticulum.

The knockout mouse model created carries a 20kb deletion, spanning the last two thirds of *C11orf53*, all of *C11orf92*, and the first third of *C11orf93*. $C11orf^{-/-}$ occur at lower frequency than expected, and show around 50% survival in the week following weaning. Tissue staining in the intestines shows reduced staining of mucins secreted by goblet cells in $C11orf^{-/-}$ mice compared to $C11orf^{+/+}$ mice. Global transcript expression analysis on intestinal tissue shows many immune system-related genes have altered expression between the genotypes, while

blood panels show that *C11orf*$^{-/-}$ mice have low red and white blood cells. However, no obvious CRC phenotype has been observed in aged mice, or when crossed onto the *Apc*$^{Min/+}$ susceptible background, or via chemical induction.

Ex-vivo analysis of the knockout mice has utilised the culture of cell lines and intestinal organoids. Cell lines showed no differences in cell proliferation or migration between the genotypes. However *C11orf*$^{-/-}$ organoids did have reduced budding compared to *C11orf*$^{+/+}$ organoids, a phenotype that is linked to defective intestinal homeostasis.

In summary, I have created a mouse model deleting the 11q23.1 CRC risk locus. Deletion of these genes has not caused CRC to develop, but there is evidence of alteration in the intestinal, immune and blood systems, and the model should greatly aid in further understanding the functional role of these genes and their precise role in CRC risk. I hypothesise that the genes play a role in the endoplasmic reticulum, and disruption of these genes alters secretion, affecting goblet cells, the Wnt signalling pathway and the immune system.

Lay Summary

Colorectal cancer (CRC) has a major health impact in the UK, where it is the fourth most common cancer and the second highest cause of cancer death. As with other cancers, there are many different factors involved in whether someone develops CRC. Some of these are related to a person's lifestyle, and some are genetic. Much work has been done in identifying and studying these genetic factors, to understand as much as possible about this disease.

A large-scale study of people with and without CRC found a particular region on chromosome 11, containing three genes, that is linked to CRC risk. People with the risk version of this region have lower activity of the three genes located there, called *C11orf53*, *C11orf92* and *C11orf93*, but the function of these genes is currently unknown, so we do not know how they are involved in CRC development.

To investigate the *C11orf* genes and their role in CRC, I looked at the location of their product in intestinal tissue, which shows they are located in specific areas of the intestine, which can help us further our understanding of their cellular function. I also saw intriguing differences in the expression of these genes in the intestines of healthy mice compared with a type of mouse known to be susceptible to CRC ($Apc^{Min/+}$).

As stated above the *C11orf* genes are expressed in mice, hence I made mice that are completely lacking the region to see what the effect of deleting the genes would be. The mice null for the three genes have altered immune system activity, and low levels of both red and white blood cells. They have a disruption in the cells that produce mucus lining their intestines. They also have difficulties in surviving weaning, with only around half making past this point. However, they do not currently show any development of CRC, even when old or given chemicals which can cause CRC. When we took intestinal tissue from the mutant mice and produced mini-guts, we saw that there was a difference to mini-guts produced from normal mice, similar to other known genetic disruptions in the intestine.

There is still a lot of work to be done to understand what these genes do and how they are involved in CRC, but my work confirms that they have a cellular role in the function of the intestines and suggests they are involved in the immune system. Hopefully, this new information and the avenues it leads to will ultimately enable an improvement in prevention of the disease and indeed impact on public health.

Declaration

I declare that this thesis was composed entirely by myself and that all research presented is my own except for where clearly indicated. No part of this work has been submitted for any other degree or qualification.

Signed

Date

Acknowledgements

I couldn't have made it through this PhD without my amazing supervisor Susan Farrington, who has helped me so much during this rollercoaster of a project, on both scientific and emotional sides. I can't imagine a better supervisor to have guided me through it! Thank you to also Mark Arends, Malcolm Dunlop and Farhat Din for your comments and advice.

Vidya Rajasekaran taught me everything I know about mouse work, did most of the IHC, and took over this project with such good grace when I needed to start writing. Anna Maria Ochocka-Fox showed me the cell activity assays and western blots, and cultured multiple organoid batches. You are both so generous with your time and knowledge, whether with helping me in the lab or discussing my results with me.

Stuart Reid for help with various PCR and cloning things, Marion Walker for making sure the lab actually functions, Alice Hardwick, Debbie Baishnab and Calum Robertson for pathology advice, Imke van Ettinger for vet advice, Maria Timofeeva and Vicky Svinti for stats advice, and all the rest of the Colon Cancer Genetics Group, and the wider C3 lab/offices – there are too many of you to list, but I feel very lucky that I got to spend four years in such a friendly, supportive environment. Thank you for all the help! And to my lunch crew, thank you for guaranteeing laughter in the middle of every day.

Special thanks to the BRF and Evans animal units, especially Gary Waugh; the Division of Pathology; IGMM Technical Services and Stores; the IGMM Advanced Imaging Resource; the IGMM Bioinformatics Service; and the University of Edinburgh Counselling Service.

Beyond the IGMM, thank you to my wonderful family and friends who have been so patient with me. Jessie Greenfield, Rosie Newton and Isla Tyrrell have all been there, ready to celebrate, encourage or distract as required. Your willingness to listen to me ramble about science completely outside your own fields is much appreciated!

And finally, my amazing parents, Jess Osborn and Jaqui Devereux, who have loved, supported and encouraged me tirelessly through not just this degree but my whole life, in more ways that I can list (but special thanks for the thesis proofreading).

List of Abbreviations

5-FU	Fluorouracil
ACF	Aberrant crypt foci
A-FAP	Attenuated familial adenomatous polyposis
AOM	Azoxymethane
APC	Adenomatous polyposis coli
APS	Ammonium persulphate
BMI	Body mass index
BMP	Bone morphogenetic protein
BrDU	Bromodeoxyuridine
CA-CRC	Colitis-associated colorectal cancer
cDNA	Complementary DNA
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CK1	Casein kinase 1
CMS	Consensus molecular subtype
COX	Cyclooxygenase
CRC	Colorectal cancer
DAB	3,3'-diaminobenzine
DLI	Distal large intestine
DM	Diabetes mellitus
DMH	Dimethylhydrazine
DNA	Deoxyribonucleic acid
DSI	Distal small intestine
DSS	Dextran sulphate sodium
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
ER	Endoplasmic reticulum

EYFP	Enhanced yellow fluorescent protein
FAP	Familial adenomatous polyposis
FCS	Fetal calf serum
FFPE	Formalin-fixed paraffin-embedded
FZD	Frizzled
GDF	Growth and differentiation factor
GO	Gene ontology
gRNA	Guide RNA
GSK3	Glycogen synthase kinase 3
GWAS	Genome-wide association study
H&E	Haematoxylin and eosin
Hb	Haemoglobin
HRP	Horseradish peroxidase
IBD	Inflammatory bowel disease
IHC	Immunohistochemistry
IMPC	International Mouse Phenotyping Consortium
IVC	Individually ventilated cage
LI	Large intestine
MAF	Minor allele frequency
MAP	*MUTYH* associated polyposis
MCHC	Mean corpuscular haemoglobin concentration
MCV	Mean corpuscular volume
MMR	Mismatch repair
mRNA	Messenger RNA
MSI	Microsatellite instability
mTFP1	Monomeric teal fluorescent protein 1
MYA	Million years ago
NBF	Neutral buffered formalin
PAM	Protospacer adjacent motif
PAS	Periodic acid Schiff
PBS	Phosphate-buffered saline

PBST	PBS-tween
PCR	Polymerase chain reaction
PCV	Packed cell volume
PLI	Proximal large intestine
PSI	Proximal small intestine
PVDF	Polyvinylidene fluoride
qRT-PCR	Quantitative real-time PCR
R/A	Rectum/anus
RBC	Red blood cell
RDW	Red blood cell distribution width
RNA	Ribonucleic acid
RONs	Reactive oxygen and nitrogen species
RT	Room temperature
SDS	Sodium dodecyl sulphate
SI	Small intestine
SNP	Single nucleotide polymorphism
SRB	Sulforhodamine B
TGF	Transforming growth factor
TNM	Tumour, node, metastasis
UC	Ulcerative colitis
UPR	Unfolded protein response
UTR	Untranslated region
UV	Ultraviolet
WBC	White blood cell
WT	Wild type

Table of Contents

Chapter 7: Discussion

Supplementary data (on disk)

Chapter 1: Introduction

1.1 Colorectal cancer

1.1.1 Colorectal tumours

Colorectal cancer (CRC) is a term for cancers of the large intestine, spanning the appendix to the rectum (figure 1.1). The colon is divided into right and left, which have different embryonic origin. The right colon, comprising appendix, caecum, ascending colon, hepatic flexure and proximal two-thirds of the transverse colon, is derived from the embryonic midgut, while the left colon, comprising distal third of the transverse colon, splenic flexure, descending colon, sigmoid colon, rectosigmoid junction, rectum and anus, is derived from the embryonic hindgut (Glebov et al., 2003). The right and left colon can show distinct patterns of gene expression.

The intestines comprises four layers: innermost mucosa, submucosa, muscularis propria and outermost serosa (Treuting and Dintzis, 2012; Treuting et al., 2012; Young et al., 2006) (figure 1.2). The mucosa itself has three layers: innermost simple columnar epithelium, lamina propria (connective tissue) and muscularis mucosae (thin smooth muscle). In both small and large intestine, the epithelium has crypt structures. The small intestine also has villi to increase surface area for nutrient absorption, while the large intestine has many secretory cells producing mucus to aid movement of faecal matter as water is absorbed. The submucosa contains blood vessels, lymphatic vessels and nerve. The muscularis has two layers of muscle, the inner layer being in rings perpendicular to the long axis of the intestine and the outer layer being longitudinal. The serosa is connective tissue.

Over 90% of colorectal carcinomas are adenocarcinomas, due to the high concentration of glands found in the tissue (Cancer Research UK, 2016). The typical progression of colorectal cancer is from aberrant crypt focus, to adenomatous polyp, to adenocarcinoma (Hamilton et al., 2000). The earliest sign of neoplasia in the intestines is aberrant crypt foci (ACF), which are characterised by dysplastic or hyperplastic crypts (Nucci et al., 1997). ACF can develop into polyps, which project out from the mucosa. Hyperplastic polyps, while showing accelerated cell proliferation, are not considered to have the same risk of malignancy as neoplastic polyps, in which the cells have begun to lose differentiation. Neoplastic polyps that originate from glandular tissue are known as adenomatous polyps, or adenomas, and

are typically removed upon detection due to their high risk of progressing to adenocarcinoma (Atkin and Saunders, 2002). The defining feature of an adenocarcinoma is invasion beyond the muscularis mucosae into the submucosa.

There are three systems for describing the stage of a CRC: tumour, node, metastasis (TNM) staging, number staging and Dukes' staging, in addition to the grade of the tumour (Cancer Research UK, 2016; Hamilton et al., 2000). The TNM system is most commonly used in the UK and is the most descriptive. TNM staging assigns a number to the growth of the tumour (T), the extent of lymph node infiltration (N) and whether or not metastasis has occurred (M). T1 cancers have invaded the submucosa, T2 cancers have reached the muscularis propria, T3 cancers have reached the serosa, and T4 cancers have grown beyond the serosa to invade other organs or structures (figure 1.3). Tis, or carcinoma in situ, refers to growths that have not penetrated beyond the mucosa. N0 cancers have not spread to any lymph nodes, N1 cancers have spread cancer cells to 1-3 nearby lymph nodes, and N2 cancers have cancer cells present in at least four lymph nodes. M0 cancers have not spread to any distant organs but M1 cancers have distant metastasis.

The Dukes' system has only four broad categories. Stage A cancers are only in the submucosa or muscularis of the colon, stage B cancers are in the serosa or beyond but have not spread, stage C cancers have reached the lymph nodes and stage D cancers have metastasised. The number system applies the same categories as the Dukes' system, but they are 1-4 rather than A-D. Additionally, stage 2 is divided into 2A and 2B, depending on whether the cancer has grown up to or beyond the serosa, and stage 3 is divided into 3A, 3B and 3C, depending on how far into the colon the tumour has grown and how many lymph nodes contain cancer cells. A comparison of how each system would categorise a given cancer is given in table 1.1.

Cancers may be given a grade along with the stage. Rather than describe the extent of the tumour growth, the grade refers to the appearance of the individual cells. Grade 1 tumours contain cells that are still largely differentiated, grade 2 tumour cells are moderately differentiated, grade 3 tumour cells are poorly differentiated, and grade 4 tumours are almost entirely undifferentiated cells.

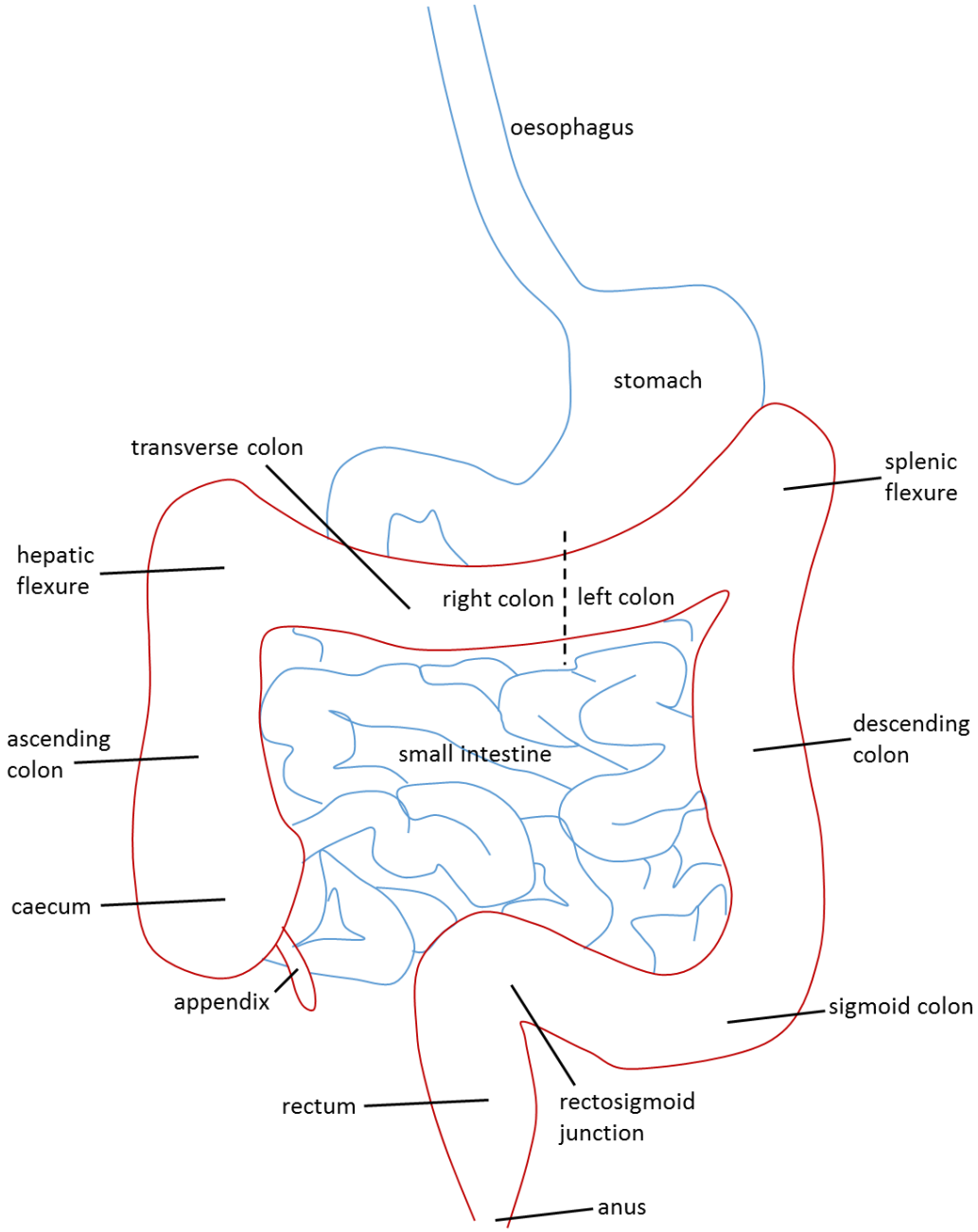


Figure 1.1: Annotated human large intestine. The appendix to the proximal two thirds of the transverse colon is collectively the right colon, and is derived from the embryonic midgut. The distal third of the transverse colon to the anus is collectively the left colon, and is derived from the embryonic hindgut.

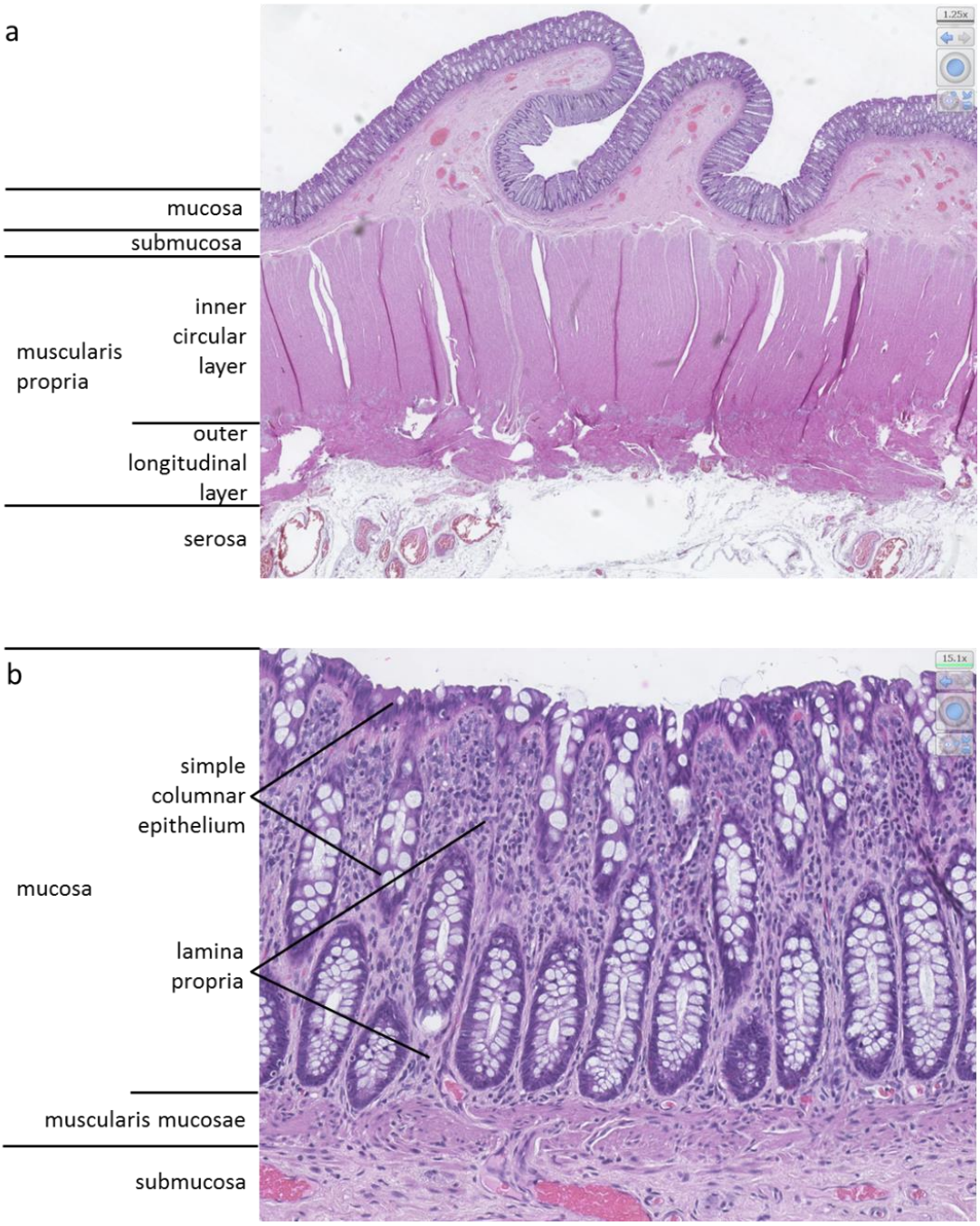


Figure 1.2: Haematoxylin and eosin stained tissue sections of human colon, taken at (a) 12.5x, showing the entire cross-section with mucosa, submucosa, muscularis propria and serosa layers, and (b) 150x, showing the epithelium, lamina propria and muscularis mucosae layers within the mucosa.

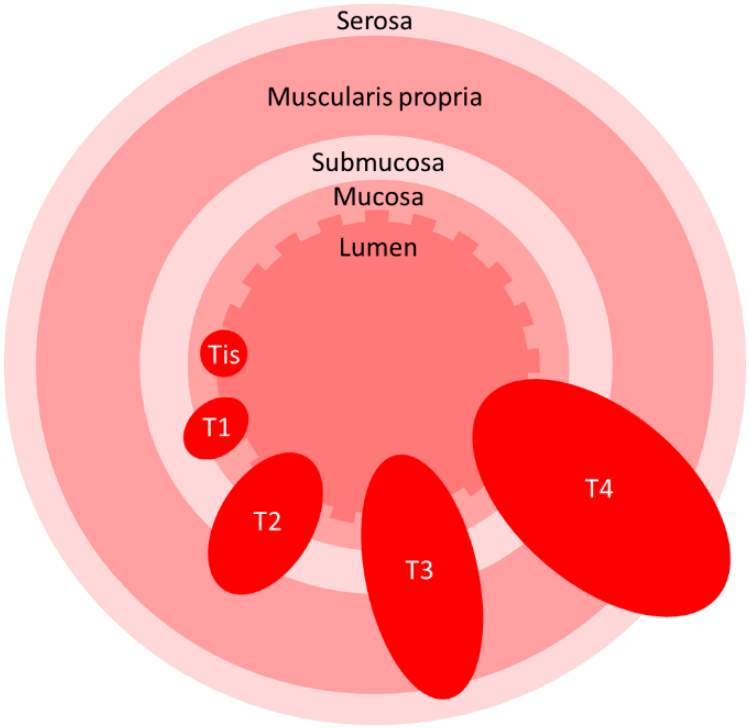


Figure 1.3: Cross section of the colon showing tissue layers and tumour stages, with tumours numbered according to the TNM staging system.

Description	TNM	Number	Dukes'
Tumour in situ present in the mucosa only	Tis		
Tumour is in the submucosa only	T1 N0 M0	1	A
Tumour has reached the muscularis propria	T2 N0 M0		
Tumour has reached the serosa	T3 N0 M0	2A	B
Tumour has grown beyond the serosa but has not spread to lymph nodes or distant organs	T4 N0 M0	2B	
Tumour is in the submucosa and has spread to 1-3 lymph nodes but no distant organs	T1 N1 M0	3A	C
Tumour has reached the muscularis and has spread to 1-3 lymph nodes but no distant organs	T2 N1 M0		
Tumour has reached the serosa and has spread to 1-3 lymph nodes but no distant organs	T3 N1 M0	3B	
Tumour has grown beyond the serosa and has spread to 1-3 lymph nodes but no distant organs	T4 N1 M0		
Tumour is any size and has spread to at least four lymph nodes but no distant organs	T* N2 M0	3C	
Tumour is any size with any degree of lymph node spreading and has metastasised to distant organs	T* N* M1	4	D

Table 1.1: Comparison of the TMN, numbers and Dukes' systems of categorising CRC stage, with a description of the cancer at each stage.

1.1.2 Incidence, treatment and mortality

In the UK, CRC is the fourth most common cancer, with 41804 new cases in 2015, amounting to 12% of all new cancer cases that year (Cancer Research UK, 2016). The more common cancers are breast cancer, prostate cancer and lung cancer. Due to the sex differences in the incidence of breast and prostate cancers, CRC is the third most common cancer in both men and women when the sexes are considered separately. Cancers of the right colon are more common in women than in men, and left colon cancers are more common in men than in women.

Treatment for CRC varies with the stage at which it is diagnosed. Surgery to remove the growth is the primary treatment for CRC and can be local excision, resection with anastomosis, resection with colostomy, or total colectomy, depending on how extensive the cancer is and how much of the colon must be removed; this should be accompanied by lymphadenectomy (Cunningham et al., 2010). The colon is highly sensitive to radiation and so radiotherapy is not routinely used in the treatment of colonic cancer, but neoadjuvant or adjuvant radiotherapy is used to treat rectal cancer (Häfner and Debus, 2016).

Chemotherapy drugs used to treat CRC include fluorouracil (5-FU), a thymidylate synthase inhibitor that blocks thymidine synthesis, leading to thymineless death in rapidly dividing cells (Longley et al., 2003); capecitabine, which is metabolised to 5-FU (Walko and Lindley, 2005); and oxaliplatin, a planar platinum-based drug which reacts with deoxyribonucleic acid (DNA) bases to create intra- and interstrand crosslinks, causing inhibition of DNA replication and transcription (Graham et al., 2004).

There were 16384 deaths due to CRC in the UK in 2016, which was 10% of all cancer deaths that year (Cancer Research UK, 2016). It is the second highest cause of cancer death in the UK, behind lung cancer, although in men alone there are more deaths from prostate cancer than CRC, and more deaths from breast cancer than CRC in women alone. England and Wales have 57% 10-year survival, 59% five-year survival and 76% one-year survival. Men have higher survival than women. When CRC is diagnosed at the earliest stage, over 90% of patients survive beyond five years, but when it is diagnosed at the latest stage, the five-year survival is below 10%.

Incidence in the UK has been stable overall over the last decade, with a 4% decrease in men and 2% increase in women (Cancer Research UK, 2016). Mortality has decreased 17% in men over the same period and 12% in women, giving 14% reduction in mortality overall.

1.1.3 Genetic changes in colorectal cancers

Colorectal cancers can be divided into subtypes, depending on the genomic instability present: those with chromosomal instability (CIN), those with microsatellite instability (MSI), and those with CpG island methylator phenotype (CIMP). These pathways are not mutually exclusive and a tumour can have characteristics of more than one. 65-85% of sporadic colorectal cancers exhibit CIN, characterised by increased rates of gain or loss of chromosome regions or whole chromosomes, leading to aneuploidy and high loss of heterozygosity (Pino and Chung, 2014). MSI is detected in around 12% of sporadic CRC cases and is a hallmark of deficiencies in DNA mismatch repair (MMR) machinery, such as *MLH1* and *PMS2* (Boland and Goel, 2010). CIMP is characterised by hypermethylation of promoter CpG islands, resulting in the inactivation of multiple tumour suppressor genes, and is present in around 20% of sporadic CRC cases (Mojarad et al., 2013). Sporadic tumours with MSI also have CIMP, as CIMP provides the mechanism for inactivating the MMR genes; cancers associated with a hereditary MMR defect such as Lynch syndrome have MSI without CIMP (Jass, 2007).

The Wnt signalling pathway is disrupted in over 90% of sporadic colorectal cancers (Cancer Genome Atlas, 2012). The canonical Wnt signalling pathway is an important regulator of cell proliferation, differentiation, adhesion and migration, and so plays key roles in embryogenesis and adult homeostasis (Logan and Nusse, 2004; MacDonald et al., 2009). It is particularly important in tissues with high cell turnover, such as the epithelium of the intestines. Binding of secreted Wnt to its Frizzled (FZD) receptor causes inhibition of the β-catenin degradation complex of adenomatous polyposis coli protein (APC), glycogen synthase kinase 3 (GSK3), casein kinase 1 (CK1) and axin, allowing β-catenin to accumulate and translocate into the nucleus where it binds TCF/LEF transcription factors and stimulates expression of Wnt-responsive genes (Behrens et al., 1998; Gregorieff and Clevers, 2005; Molenaar et al., 1996; Schatoff et al., 2017) (figure 1.4).

The most common mutation in the Wnt signalling pathway is biallelic inactivation of *APC*, member of the β-catenin destruction complex; *APC* mutations are seen in over 80% of sporadic colorectal cancers, and are one of the earliest identified mutations in the CRC pathway (Cancer Genome Atlas, 2012; Powell et al., 1992). The *APC* gene is on chromosome arm 5q, which has high rates of loss through CIN (Al-Sohaily et al., 2012). Other frequent mutations found in colorectal cancers that affect the Wnt signalling pathway include activating mutations in β-catenin gene *CTNNB1*, upregulation of the Wnt receptor *FZD*, and inactivation of *DKK-1*, which encodes the DICKKOPF-1 competing ligand for the LPR6 Wnt receptor (González-Sancho et al., 2005; Morin et al., 1997; Terasaki et al., 2002).

The RAS/MAPK pathway is a cell-signalling pathway which communicates extracellular signals from the cell membrane to the nucleus, in particular regulating cell proliferation by transducing signals from growth factors and proto-oncogenes (Fang and Richardson, 2005; Zhang et al., 2002). Around 60% of sporadic colorectal cancers have mutations in the RAS/MAPK pathway, with 55% having mutually exclusive activating mutations in either *KRAS*, *NRAS* or *BRAF* (Cancer Genome Atlas, 2012). Extracellular signals promote guanosine diphosphate binding to Ras proteins, including K-ras and N-ras, leading to activation of Raf proteins such as B-raf, and triggering a kinase cascade culminating in the phosphorylation of transcription factors (Fang and Richardson, 2005; Zhang et al., 2002).

TP53, encoding the tumour suppressor p53, is mutated in around 60% of sporadic adenocarcinomas, but not in adenomas, meaning *TP53* mutations occur relatively late in colorectal tumourigenesis (Vogelstein et al., 1988). p53 has multiple roles in preserving genome integrity including arresting the cell cycle at the G1 checkpoint, activating genes encoding DNA repair machinery, and triggering apoptosis in cells with excessively damaged DNA or that are at risk of aberrant mitosis (Kastenhuber and Lowe, 2017). It is therefore a highly important tumour suppressor gene and is the most widely mutated gene in cancer (Kandoth et al., 2013).

The transforming growth factor (TGF) β pathway plays roles in a variety of cell processes, including proliferation, apoptosis, stem cell renewal, differentiation and adult tissue homeostasis (Shi and Massagué, 2003; Weiss and Attisano, 2013). The TGF-β superfamily of secreted signalling ligands includes the bone morphogenetic protein (BMP), Nodal, Activin and growth and differentiation factor (GDF) families. TGF-β ligands bind to receptor serine/threonine kinases at the cell surface and the signal is transduced to the nucleus via

Smad proteins. Many components of the TGF-β signalling pathway are found to be mutated in CRC, including *TGFBR1*, *ACVR2A*, *SMAD3* and *SMAD4* (Cancer Genome Atlas, 2012).

Recently, analysis of gene expression in a large dataset of over 4000 patients was used to define four consensus molecular subtypes (CMS) of CRC (Guinney et al., 2015). CMS1, or MSI immune, represents 14% of tumours and is defined as being MSI, CIMP-high and hypermutated, with *BRAF* mutations, and immune infiltration and activation. CMS2, called the canonical subgroup, comprises 37% of tumours and is CIN-high with WNT and MYC activation. 13% of tumours are CMS3, or metabolic, meaning they are mixed MSI status, CIN-high and CIMP-low, with *KRAS* mutations and metabolic deregulations. CMS4, or mesenchymal, represents 23% of tumours and is CIN-high, with stromal infiltration, TGF-β activation and angiogenesis. The remaining 13% show a mixture of features, and may be a transitional phenotype, or may represent intratumour heterogeneity. These subtypes, which were identified from patient samples, have been shown to also be present in in vitro and in vivo CRC models (Linnekamp et al., 2018). The use of CMS offers the potential for improved precision in CRC treatment (Dienstmann et al., 2017).

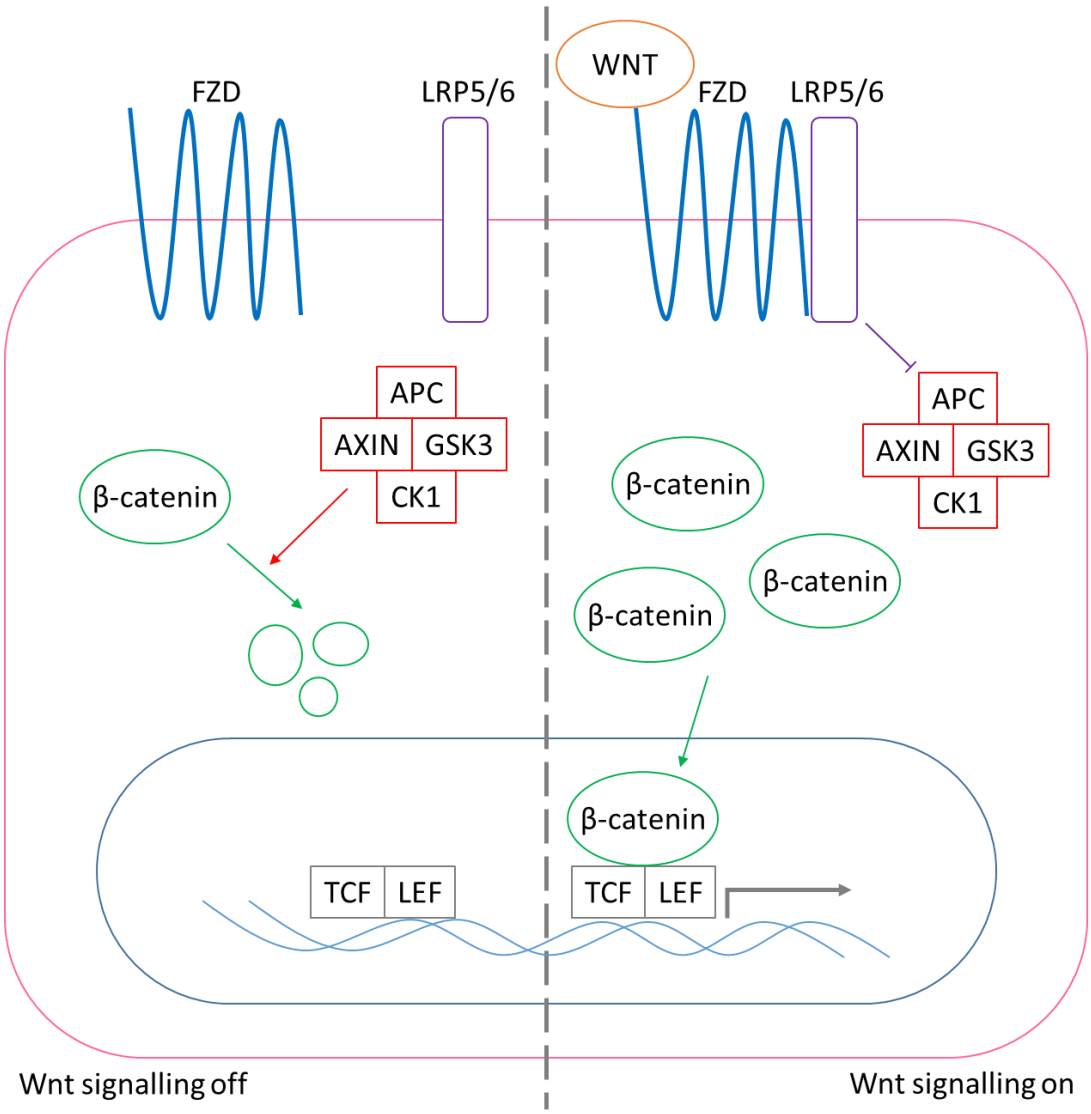


Figure 1.4: Schematic of basic canonical Wnt signalling. When Wnt signalling is inactive, GSK3, in complex with APC/AXIN/CK1, phosphorylates cytoplasmic β-catenin and so targets it for degradation. When Wnt signalling is activated, secreted Wnt binds to the transmembrane receptors Frizzled and LRP5/6. GSK3 is then phosphorylated, inhibiting its own activity. β-catenin accumulates in the cytoplasm and is transported into the nucleus, where it binds to TCF/LEF transcription factors and promotes transcription of Wnt-responsive genes.

1.1.4 Genetic risk factors for CRC

The known genetic factors which contribute to CRC susceptibility can be broadly split into two groups: high penetrance, low frequency mutations, which are only present in a small percentage of the population but which confer a greatly increased risk of CRC development; and low penetrance, high frequency variants, which are more common but carry a lower associated risk (Peters et al., 2015) (figure 1.5). High penetrance mutations manifest in hereditary syndromes and so have largely been identified via studies of affected families. The three most common hereditary syndromes are familial adenomatous polyposis, Lynch syndrome and *MUTYH* associated polyposis.

Familial adenomatous polyposis (FAP) is an autosomal dominant syndrome characterised by the development of hundreds of adenomatous polyps, mostly in the left colon, which are detectable between ages 10 and 20 (Galiatsatos and Foulkes, 2006; Talbot et al., 2000). These adenomas have a high tendency to progress to adenocarcinomas, giving almost 100% incidence of CRC by age 40. Around 1% of all CRC cases is due to FAP. FAP is caused by an inherited inactivating mutation in one copy of *APC*, the gene so frequently mutated at the initiation of sporadic CRC. There are many different mutations observed in *APC*, with the site of the mutation correlating with the severity of FAP (Nieuwenhuis and Vasen, 2007), such as mutations before codon 157 or after codon 1595 (of a total 2843 codons) being associated with attenuated FAP (A-FAP). A-FAP adenomas typically number below 100 and are more proximal than in classic FAP, and CRC onset in A-FAP occurs on average 15 years later than classic FAP (Galiatsatos and Foulkes, 2006).

Lynch syndrome, also known as hereditary non-polyposis colorectal cancer, is an autosomal dominant syndrome that accounts for around 3% of CRC cases, with the mean age of cancer diagnosis being 45 (Lynch et al., 2009; Peltomaki et al., 2000). The majority of adenomas develop in the right colon and show accelerated carcinogenesis, progressing from adenoma to adenocarcinoma more rapidly than non-Lynch syndrome cancers (Rijcken et al., 2002). Lynch syndrome is caused by germline mutations in MMR genes, with *MLH1*, *MSH2*, *MSH6*, *PMS2*, *PMS1* and *MLH3* all being linked to Lynch syndrome heritability, although *MLH1*, *MSH2* and *MSH6* account for over 95% of known mutations associated with Lynch syndrome (Lynch et al., 2009; Peltomäki, 2005). Lynch syndrome tumours are therefore defined by high MSI, resulting from the defective DNA repair machinery.

MUTYH associated polyposis (MAP) is an autosomal recessive syndrome, caused by a biallelic mutation in the *MUTYH* gene (Farrington et al., 2005; Mork and Vilar, 2016; Poulsen and Bisgaard, 2008). It has variable presentation, ranging from no polyps to polyp numbers almost equal to classic FAP. MUTYH is a DNA glycosylase involved in base excision repair for oxidative damage, and mutation of *MUTYH* results in G:C > A:T transitions. MAP tumours have high incidence of such transitions occurring within the *APC* gene (Sieber et al., 2003).

In contrast to high penetrance mutations, low penetrance variants do not typically have a strong enough effect to create a clear familial syndrome and so contribute to "sporadic" risk, although they can influence a family history of CRC. These variants are identified via genome wide association studies (GWAS). Identified variants of this type include *CTNNB1* and *CDH1*, both members of the Wnt signalling pathway; *TGFB1*, *BMP2* and *BMP4*, members of the TGF-β signalling pathway; and *POLD3*, a DNA polymerase with proofreading ability (Dunlop et al., 2016; Peters et al., 2015).

Due to their increased frequency, an individual may carry risk alleles for multiple low penetrance variants, the effects of which combine into their overall genetic risk (Dunlop et al., 2012). Common variants may also be present in a person carrying a high-effect mutation and so further modify their risk. As many low penetrance variants are overrepresented in adenomas as well as adenocarcinomas, it has been suggested that such variants affect CRC risk by modifying susceptibility to the precursor adenomas rather than the rate at which adenomas progress (Carvajal-Carmona et al., 2013).

Theoretically there are also variants of intermediate effect and frequency, but as they do not have either the effect size to present in family studies, or the frequency to be detectable in GWAS, they have so far been very difficult to identify; this will likely change as the cost of genome sequencing continues to drop and GWAS can be carried out on large enough cohorts to provide the power to find these intermediate variants.

Comparing the genes which are the cause of hereditary CRC syndromes, genes which have been identified as low penetrance CRC variants, and genes which frequently accumulate mutations during CRC development, there are many overlaps in the pathways implicated, such as the Wnt and TGF-β/BMP signalling pathways.

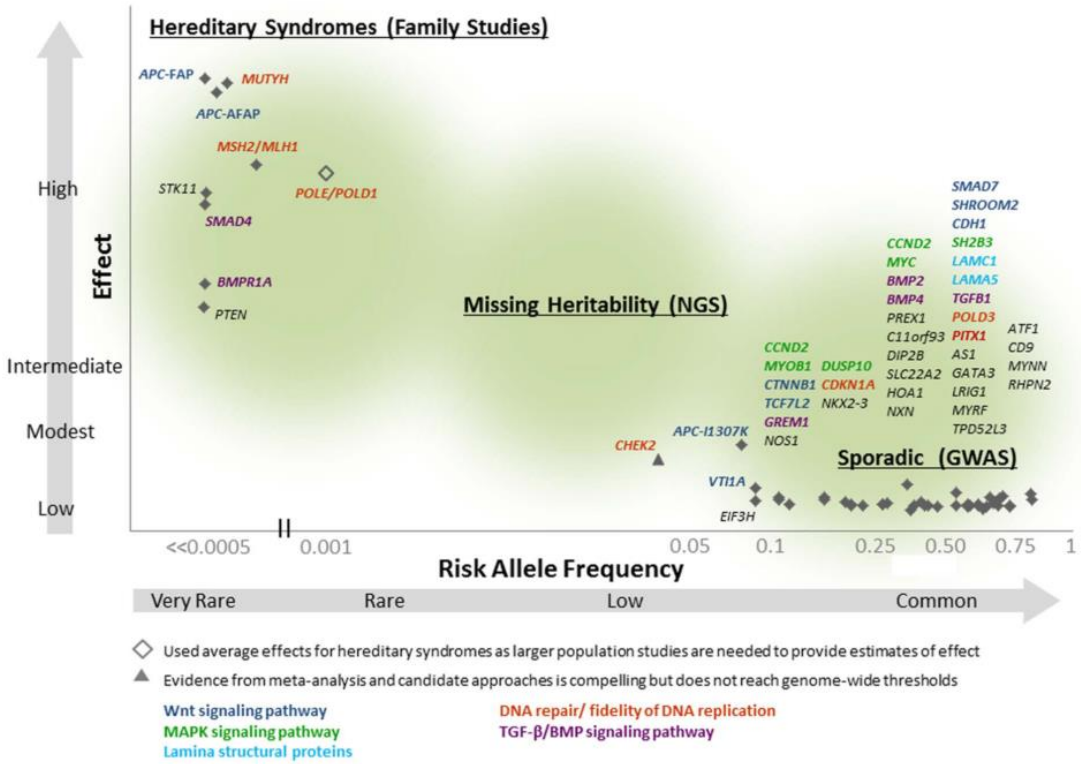


Figure 1.5: Known CRC susceptibility loci arranged by risk allele frequency and effect size. High effect, low frequency variants cause strong hereditary syndromes and so have largely been identified by familial studies, while low effect, high frequency variants are typically identified by GWAS studies of sporadic cancers. The hypothesised "missing heritability" genes have too weak an effect to be detectable in familial studies, but are not common enough to be detectable in GWAS. The continued development of next generation sequencing may aid identification of such genes. Figure taken from Peters et al, 2015.

1.1.5 Environmental risk factors for CRC

1.1.5.1 Medical conditions

Inflammatory bowel disease (IBD), mainly comprising Crohn's disease and ulcerative colitis (UC), is characterised by chronic inflammation. The risk of CRC is 70% higher in the IBD population than the non-IBD population, with an individual's risk increase correlating with duration of disease, extent of colitis and degree of inflammation (Canavan et al., 2006; Castaño-Milla et al., 2014; Lutgens et al., 2013). While IBD underlies only 1-2% of CRC cases, CRC is the cause of 15% of deaths in IBD patients (Munkholm, 2003), and IBD is the third highest risk factor for CRC, behind FAP and Lynch syndrome (Kulaylat and Dayton, 2010).

Overall CIN and MSI occur at the same rate in colitis-associated CRC (CA-CRC) as in sporadic CRC, but the timing and frequency of specific mutations can show different patterns (Lakatos and Lakatos, 2008). Rather than being an initiating mutation as in sporadic CRC, *APC* mutations occur later in the progression of CA-CRC and so at a lower frequency (Aust et al., 2002). Conversely, *TP53* mutations occur at a higher rate in CA-CRC as they are earlier events than in sporadic CRC (Hussain et al., 2000).

The pathogenesis of CA-CRC is complex and there are many mechanisms that contribute. Many mucosal inflammatory mediators are induced by the chronic inflammation that characterises IBD, including inflammatory cytokines, chemokines and cyclooxygenase (COX) enzymes, and some of these are linked to CRC. For example COX-2, which is induced by inflammatory stimuli including IFN-γ and TNF-α, is upregulated in colorectal cancers, and expression of COX-2 correlates with tumour stage and depth of invasion (Elzagheid et al., 2013; Lim et al., 2010). Another pathway involved in the pathogenesis of CA-CRC is that of reactive oxygen and nitrogen species (RONs). These are produced in high amounts by inflammatory cells such as activated neutrophils and macrophages, leading to inflamed tissue having higher concentrations of RONs (Kimura et al., 1998; Rachmilewitz et al., 1995). Such species can cause denaturation of proteins and DNA alterations including base modifications, double base lesions and single- and double-strand breaks (Klaunig et al., 1998); and oxidative stress can also inactivate the MMR system (Chang et al., 2002).

Diabetes mellitus (DM) is characterised by defects in insulin secretion and/or action; type 1 DM has absolute insulin deficiency due to autoimmune destruction of insulin-secreting β-cells in the pancreas, while type 2 DM has insulin resistance. Meta-analyses have linked type

2 DM only with CRC, with 22-30% increase in CRC risk in type 2 DM patients (Jiang et al., 2011; Larsson et al., 2005; Luo et al., 2012). There are many shared risk factors between CRC and type 2 DM, such as obesity, but other possible mechanisms for this link include hyperinsulinemia due to resistance, as levels of insulin and insulin-like growth factor 1 have been shown to correlate with the number and classification of colorectal adenomas (Schoen et al., 2005).

1.1.5.2 Lifestyle factors

It is estimated that 54% of UK CRC cases in 2015 were preventable (Cancer Research UK, 2016). Key lifestyle factors that affect cancer risk include diet, physical activity, weight, alcohol consumption and tobacco smoking.

High intake of both red and processed meat have been linked to increased CRC risk by multiple meta-analyses of human studies, showing that 100-120g red meat per day increases CRC risk 17-30%, and 25-50g of processed meat per day increases risk 9-50% (Aune et al., 2013; Chan et al., 2011; Norat et al., 2002). It has also been shown that beef and blood sausage promote colon carcinogenesis in rats treated with a carcinogen (Pierre et al., 2004). In this study, a control group given no meat but supplemented with haem to match the beef diet haem content showed carcinogenesis equal to the beef diet. Haem, which is present in high concentrations in meat, stimulates production of N-nitroso compounds, leading to DNA damage (Joosen et al., 2009). 13% of UK CRC cases are due to processed meat consumption (Brown et al., 2018). In contrast, meta-analyses show high consumption of fibre reduces CRC risk, with a 10% reduction in CRC risk per 10g/day total dietary fibre and 20% risk reduction per 90g/day whole grains (Aune et al., 2011; Ben et al., 2014; Song et al., 2015). 28% of UK CRC cases are due to insufficient fibre consumption (Brown et al., 2018).

CRC risk is increased with alcohol consumption. Risk increases 17% with daily consumption of 1.5-6 units of alcohol (12.5-50g), and is 33% higher in people who drink over 6 units per day, according to one meta-analysis (Bagnardi et al., 2015). Another meta-analysis found risk increases 7% with each additional unit of alcohol drunk per day (Fedirko et al., 2011). The first metabolite of alcohol, acetaldehyde, is carcinogenic and causes DNA damage including sister-chromatid exchange and intra-strand crosslinks (Matsuda et al., 1998; Obe et al., 1986). 6% of UK CRC cases are due to alcohol consumption (Brown et al., 2018).

Physical inactivity increases colon cancer risk up to 24%, while sedentary behaviour increases colon cancer risk up to 27%; rectal cancer is not associated with physical activity, but sedentary behaviour increases rectal cancer risk up to 6% (Boyle et al., 2012; Cong et al., 2014; Robsahm et al., 2013). Sedentary behaviour, defined as waking behaviour with low energy expenditure while sitting or lying down, is not interchangeable with physical inactivity, defined as performing insufficient moderate- and vigorous-intensity activity, hence the risk being conferred individually by each being different (van der Ploeg and Hillsdon, 2017). Proposed mechanisms for the protective effect of physical activity on colorectal cancer include altered levels of hormones such as insulin and prostaglandins, increased immune function, and decreased gastrointestinal transit time (Friedenreich, 2011; McTiernan, 2008). 5% of UK CRC cases are due to too little physical activity (Brown et al., 2018).

Being overweight (body mass index (BMI) of 25-29.9) or obese (BMI of 30+) has been linked by meta-analyses to CRC risk, independent of diet and physical activity (Chen et al., 2015; Ning et al., 2010; Xue et al., 2017). Overweight men have 18% higher risk of colon cancer and 6% higher risk of rectal cancer, while obese men have 48% higher risk of colon cancer and 25% higher risk of rectal cancer. CRC risk in men increases 10% per 5 BMI units gained during adulthood. There is a 12% increase in risk of colon cancer in obese women and no association between rectal cancer and BMI in women, or between CRC and adult weight gain in women. Proposed mechanisms for the increased risk of CRC with increased weight include insulin resistance, chronic inflammation and altered adipokine secretion, with adipose tissue dysfunction as a key underlying cause (Van Kruijsdijk et al., 2009; Martinez-Useros and Garcia-Foncillas, 2016). 11% of UK CRC cases are due to overweight and obesity (Brown et al., 2018).

People who are current or former smokers have 17-25% higher risk of CRC than people who have never smoked, with risk being higher in those who began smoking at a younger age and increasing 7-11% per 10 cigarettes smoked daily (Huxley et al., 2009; Liang et al., 2009; Tsoi et al., 2009). The association may be stronger for rectal cancer than colonic cancer. There are over 60 established carcinogens present in cigarette smoke, which act through a range of mechanisms (Alavanja et al., 2004). 7% of UK CRC cases are due to tobacco smoking (Brown et al., 2018).

1.2 The 11q23.1 locus

1.2.1 Identification as a CRC risk locus

The 11q23.1 locus was identified as a CRC risk locus in a 2008 GWAS (Tenesa et al., 2008). In phase 1 of this GWAS, 1012 early-onset Scottish CRC cases and 1012 controls were genotyped for 555510 single nucleotide polymorphisms (SNPs). The highest-ranked 15008 SNPS were taken into phase 2 and genotyped in 2057 Scottish CRC cases and 2111 controls. Finally, 14500 CRC cases and 13294 controls from seven populations were genotyped for the five highest-ranked SNPs from phases 1 and 2, the populations being Scottish, English, German, Israeli, Spanish, Canadian and Japanese.

In addition to replicating known risk loci at 8q24 and 18q21, this study identified a novel risk SNP, rs3802842, at locus 11q23.1, with odds ratio=1.1 and P=5.8E-10. rs3802842 has a minor allele frequency (MAF) of 0.29. The association between this SNP and CRC risk has been replicated in multiple populations, including Austrian, African American, Dutch, Japanese, Finnish and Chinese (Hofer et al., 2017; Kupfer et al., 2010; Middeldorp et al., 2009; Tanikawa et al., 2018; Tanskanen et al., 2018; Xiong et al., 2010). There are site-specific differences in the risk conferred by rs3802842, with the risk of rectal cancer being higher than colonic cancer. The extent of this difference varies between populations.

rs3802842 lies within an intronic region and so is not a coding variant (figure 1.6). Expression analysis on colonic tissue and CRC cell lines has shown that rs3802842 genotype correlates with expression of three local genes, *C11orf53*, *C11orf92* and *C11orf93*, with expression of all three genes decreasing with increased risk alleles of the SNP (Biancolella et al., 2014; Closa et al., 2014; Smillie, 2015) (figure 1.7). This relationship does not hold for the next nearest gene, *POU2AF1*.

C11orf53 is a protein-coding gene on the forward strand with five exons. There are three known transcripts (table 1.2a). *C11orf92*, also called *COLCA1*, is believed to be a non-coding gene. It is on the reverse strand and has five exons, with five known transcripts (table 1.2b). *C11orf93*, also called *COLCA2*, is a protein-coding gene on the forward strand, which overlaps with *C11orf92*. There are nine exons and seven known transcripts (table 1.2c). In this thesis I will refer to these genes only by their *C11orf* name.

It has been proposed that the 11q23.1 locus acts as a CRC risk modifier in patients with Lynch syndrome as well as in the general population. This has been shown in a Lynch syndrome cohort with a mix of *MLH1*, *MSH2* and *MSH6* mutations (Wijnen et al., 2009) and the association with *MLH1* has been replicated, although not the association with *MSH2* or *MSH6* (Talseth-Palmer et al., 2010). The extent of this effect and its clinic usefulness has been disputed (Houlle et al., 2011; Win et al., 2013). A study looking specifically at *PMS2*-associated Lynch syndrome did not find any risk modification from 11q23.1 (ten Broeke et al., 2017). It has also been shown that 11q23.1 does not modify the extent of polyposis in FAP patients (Cheng et al., 2015),

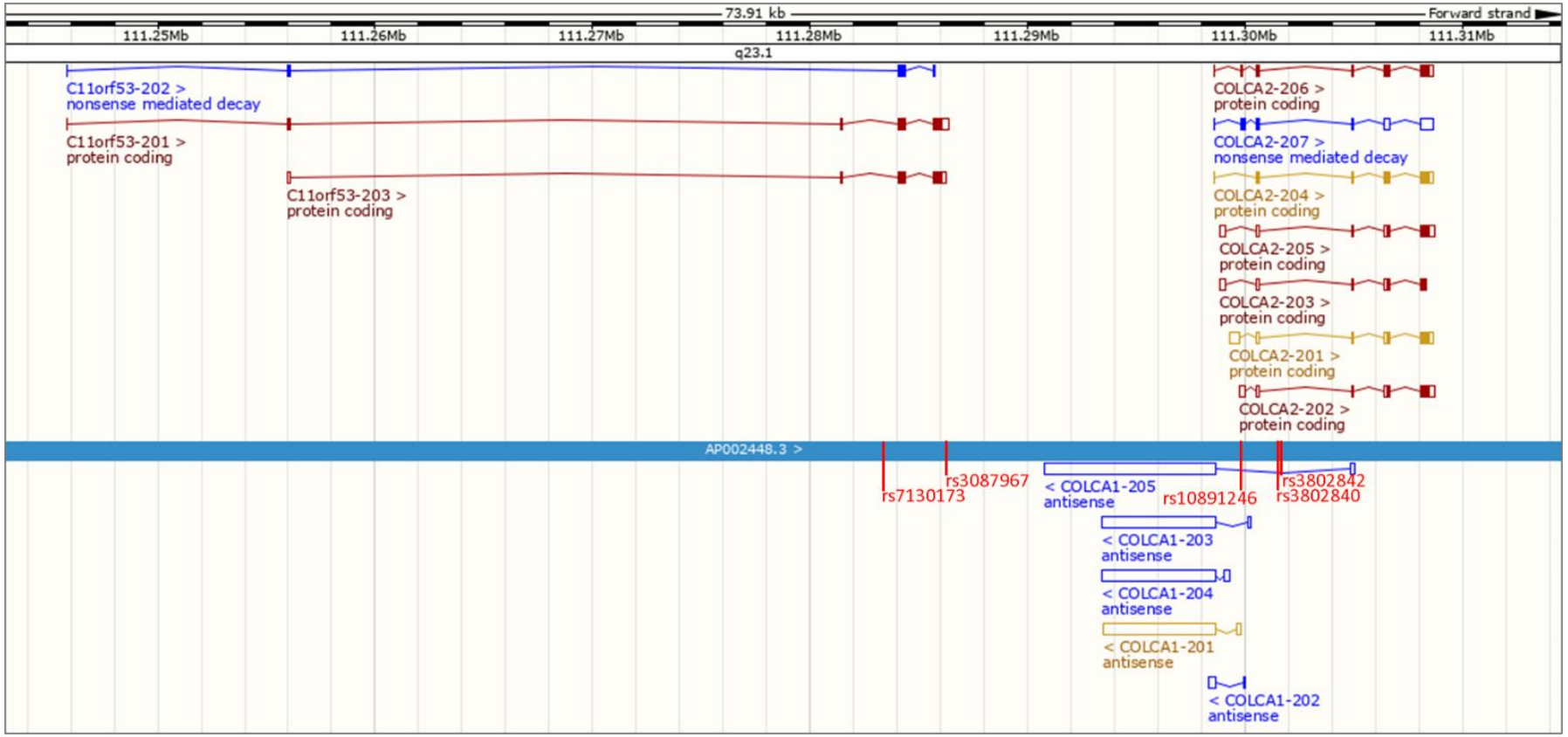


Figure 1.6: A schematic of the human locus 11q23.1, containing the genes *C11orf53*, *C11orf92* (aka *COLCA1*) and *C11orf93* (aka *COLCA2*), showing their exon structure and relative positions and sizes. Locations of key SNPs, including the original tagging SNP rs3802842 and probable causative SNP rs7130173, are marked. Image from Ensembl, July 2018.

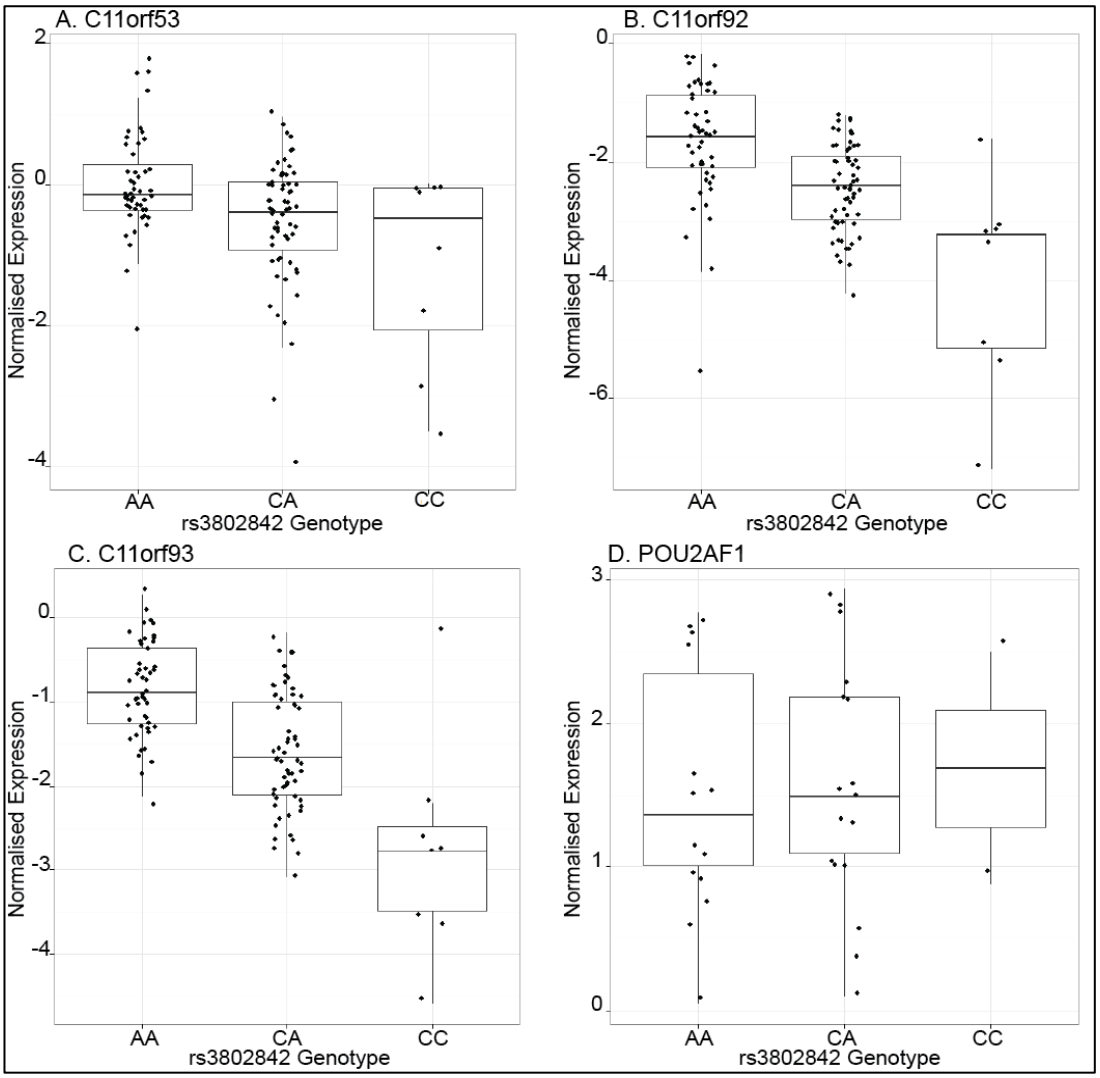


Figure 1.7: Correlation of expression of (a) *C11orf53*, (b) *C11orf92*, (c) *C11orf93* and (d) *POU2AF1* with increasing risk alleles of rs3802842. Expression data was generated using qRT-PCR from 116 colorectal tissue samples. The association between rs3802842 genotype and gene expression is significant for the three *C11orf* genes but not for *POU2AF1*. Image from Smillie, 2015.

1.2a) *C11orf53* known transcripts							
Name	**Transcript ID**	**Basepairs**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
C11orf53-203	ENST00000637637.1	1049	236aa	Protein coding	CCDS31674	Q8IXP5	NM_198498 NP_940900
C11orf53-201	ENST00000280325.5	1208	286aa	Protein coding	-	A0A1C7CYV6	-
C11orf53-202	ENST00000635886.1	545	110aa	Nonsense mediated decay	-	A0A1B0GU63	-
1.2b) *C11orf92* known transcripts							
Name	**Transcript ID**	**Basepairs**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
COLCA1-205	ENST00000620864.1	8098	No protein	Antisense	-	NM_001302644	COLCA1-205
COLCA1-204	ENST00000540738.3	5537	No protein	Antisense	-	NM_001302646	COLCA1-204
COLCA1-201	ENST00000355430.4	5442	No protein	Antisense	-	NM_207429	COLCA1-201
COLCA1-203	ENST00000532918.4	5366	No protein	Antisense	-	NM_001302645	COLCA1-203
COLCA1-202	ENST00000526150.1	391	No protein	Antisense	-	-	COLCA1-202
1.2c) *C11orf93* known transcripts							
Name	**Transcript ID**	**Basepairs**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
COLCA2-201	ENST00000398035.6	1414	154aa	Protein coding	CCDS44728	A8K830	NM_001136105 NP_001129577
COLCA2-202	ENST00000526216.1	1332	154aa	Protein coding	CCDS44728	A8K830	-
COLCA2-205	ENST00000614153.4	1264	154aa	Protein coding	CCDS44728	A8K830	NM_001271457 NP_001258386
COLCA2-204	ENST00000610738.5	975	251aa	Protein coding	CCDS73378	A8K830	NM_001271458 NP_001258387
COLCA2-206	ENST00000638573.1	1080	286aa	Protein coding	-	A8K830	-
COLCA2-203	ENST00000528846.5	823	106aa	Protein coding	-	A0A0A0MTE5	-
COLCA2-207	ENST00000639470.1	1207	133aa	Nonsense mediated decay	-	A0A1W2PQW7	-

Table 1.2: Details of the known transcripts of (a) *C11orf53*, (b) *C11orf92* and (c) *C11orf93*. Data from Ensembl, July 2018.

1.2.2 Functional understanding of 11q23.1

The region has been fine-mapped to examine SNPs in the vicinity that were not included in the original analysis (Biancolella et al., 2014; Closa et al., 2014; Smillie, 2015). Three other nearby SNPs, rs7130173 (MAF=0.29), rs3087967 (MAF=0.27) and rs10891246 (MAF=0.27), have higher correlation with expression of *C11orf53*, *C11orf92* and *C11orf93* (figure 1.8). The relationship is in the same direction as rs3802842, with the minor allele being associated with lower expression. While rs3802842 is the tagging SNP, one of these other higher-ranked SNPs may be the causal SNP for the differential expression that is observed. These SNPs are also all non-coding (figure 1.6).

While the association between CRC, rs3802842, and *C11orf53*, *C11orf92* and *C11orf93* has been widely replicated, understanding of the underlying mechanism is still greatly lacking; as the function of these genes is unknown, we cannot hypothesise how reduction of this function can lead to CRC development.

Claire Smillie carried out functional analysis of the three genes, including tissue specific expression analysis using FANTOM5 data, which showed that *C11orf53* is primarily expressed in colorectal tissue, while *C11orf92* and *C11orf93* have broader expression including male reproductive tissues, kidney and lung (Smillie, 2015). Preliminary localisation studies showed C11ORF53 may co-localise with the endoplasmic reticulum (ER) protein Sec61, and C11ORF93 has a speckled appearance within the cell, possibly also within the ER. Claire Smillie also performed extensive gene ontology enrichment analysis on gene sets derived from correlative expression within normal colorectal tissue and global expression changes in cells during overexpression or small interfering ribonucleic acid (RNA) knockdown of the *C11orf* genes. Enriched terms indicated involvement in RNA metabolic processes, the cell cycle, vesicles, and organelle membranes. The ER was also frequently arising: ER membrane correlates to *C11orf93* expression in colonic tissue, and ER membrane, ER components and ER stress response are enriched in genes altered by *C11orf53* overexpression, *C11orf53* knockdown and *C11orf92* knockdown.

The ER and unfolded protein response (UPR) have been linked to CRC and intestinal homeostasis (Drake et al., 2015; Heijmans et al., 2013; Niederreiter et al., 2013). This is in part due to the role of the ER in the correct function of goblet cells, the secretory cells of the intestines that produce large amounts of mucus and so have extensive ER to ensure correct assembly of the secreted mucus components (Hasnain et al., 2013; Heazlewood et al., 2008;

Park et al., 2009; Zhao et al., 2010). Goblet cells are part of the intestinal immune system and mucus aids prevention of inflammation by minimising intestinal epithelium contact with rough material and microorganisms; disrupted mucus production is therefore associated with inflammation and colitis (Bergstrom et al., 2015; Johansson et al., 2014; Van der Sluis et al., 2006). As such, ER stress and UPR is also associated with colitis (Hino et al., 2014; Kaser et al., 2008; Shkoda et al., 2007; Zhang et al., 2015), which is a significant risk factor for CRC, as discussed above.

An additional study shows that C11ORF93 is found in the cytoplasm of normal epithelial, immune and tumour cells within the colon, while C11ORF92 forms granules within many immune cell types including neutrophils and macrophages, and so concludes that these genes are involved in immune pathways (Peltekova et al., 2014). However, our tests of the antibodies used demonstrate a lack of the specificity desired for localisation studies, in addition to *C11orf92* having no confirmed protein to detect immunologically. More work is therefore needed on the localisation of these gene products, utilising improved tools.

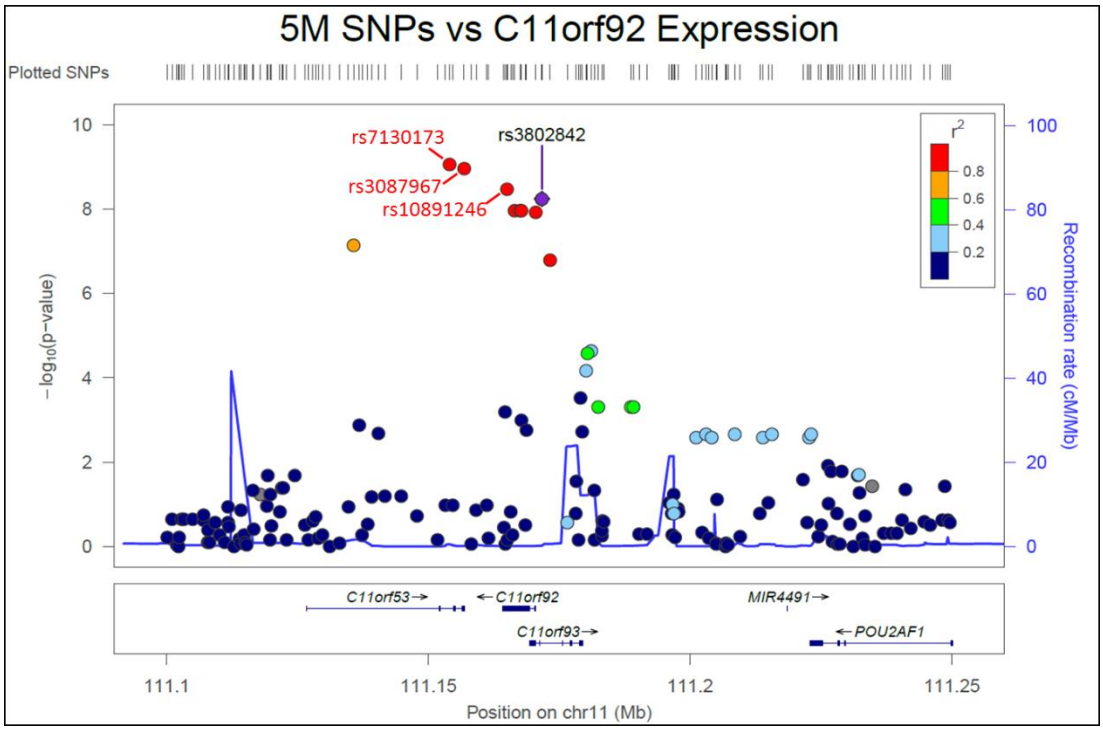


Figure 1.8: Correlation of expression of *C11orf92* with 167 SNPs from the 11q23.1 locus, using colonic gene expression in 117 people. Each dot represents one SNP, colour coded for r^2 with tagging SNP rs3802842. rs7130173, rs3087967 and rs10891246 have higher correlation with *C11orf92* expression than rs3802842. Image from Smillie, 2015.

1.2.3 Evolution and conservation in other species

1.2.3.1 Evolution of the locus

C11orf53 is the oldest of the three genes in the 11q23.1 locus, being present in almost all descendants of bony fish, but not cartilaginous fish, and so arising between 530 and 435.3 million years ago (MYA) (figure 1.9).

C11orf93 arose later, between 435.3 and 413.0 MYA, and is present only in the descendants of lobe-finned fish and not ray-finned fish (figure 1.9). As in humans, *C11orf93* orthologues are adjacent to and in the same orientation as *C11orf53* orthologues.

C11orf92 is the youngest of the three genes, being present only in placental mammals and arising between 158.6 and 105.5 MYA (figure 1.9). While *C11orf92* has no orthologues listed in genome browsers, examination of the equivalent locus (as defined by the presence of *C11orf53* and *C11orf93* orthologues) show an orthologous gene, between *C11orf53* and *C11orf93* and in the opposite orientation, in animals including armadillos, goats, dolphins, and various mouse and primate species.

There are many placental mammals that do not contain a *C11orf92* orthologue, such as pigs, rabbits, rats and gorillas. It is possible that the gene has been lost many times. However, due to the gene being possibly non-coding and quite small – only 342bp in orangutans – I believe it is also likely the orthologues of *C11orf92* have simply not yet been annotated as genes in many species, particularly those that have been less intensively studied and still only have genome scaffolds rather than full genome sequences. This hypothesis opens the prospect that not only is the gene more frequent among placental mammals, but that it may in fact be older and present in lineages beyond placental mammals.

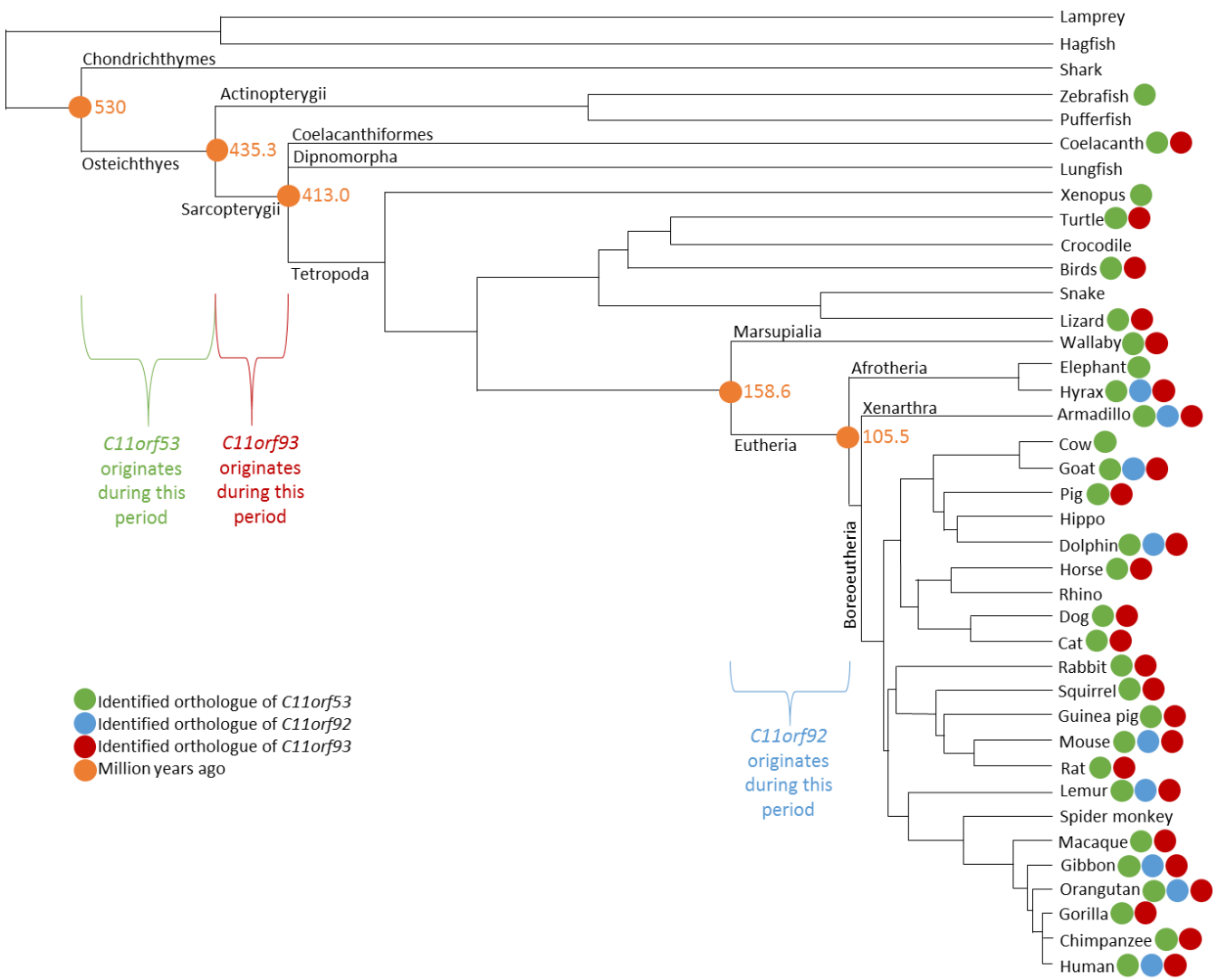


Figure 1.9: Vertebrate phylogenetic tree, with presence of orthologues of *C11orf53*, *C11or92* and *C11orf93* marked. Periods during which each gene originated are indicated, with the time of key branchpoints marked (data from Ensembl, July 2018).

1.2.3.2 Mouse orthologues

The orthologous locus in *Mus musculus* is on chromosome 9 at position A5.3. As in the human locus, the *C11orf92* orthologue is between the other two genes and in the reverse orientation, with some overlap between the 5' end of *C11orf92* and the 5' end of *C11orf93* (figure 1.10).

The mouse orthologue of *C11orf53* is *1810046K07Rik*. It is a protein coding gene with 6 exons and two known transcripts (table 1.3a). There is 82% identity of complementary DNA (cDNA) sequence and 80% identity of protein sequence between the human and mouse genes.

The mouse orthologue of *C11orf92* is *2010007H06Rik*. Like *C11orf92*, it is believed to be non-coding. There is 56% identity of cDNA sequence between human and mouse genes. There are two exons and two known transcripts (table 1.3b).

The mouse orthologue of *C11orf93* is *Gm684*, recently renamed *Colca2*. It has five exons and two known transcripts (table 1.3c). There is 73% cDNA sequence identity and 63% protein sequence identity between the human and mouse genes.

For clarity, in this thesis I will refer to the mouse genes by the names of their human orthologues.

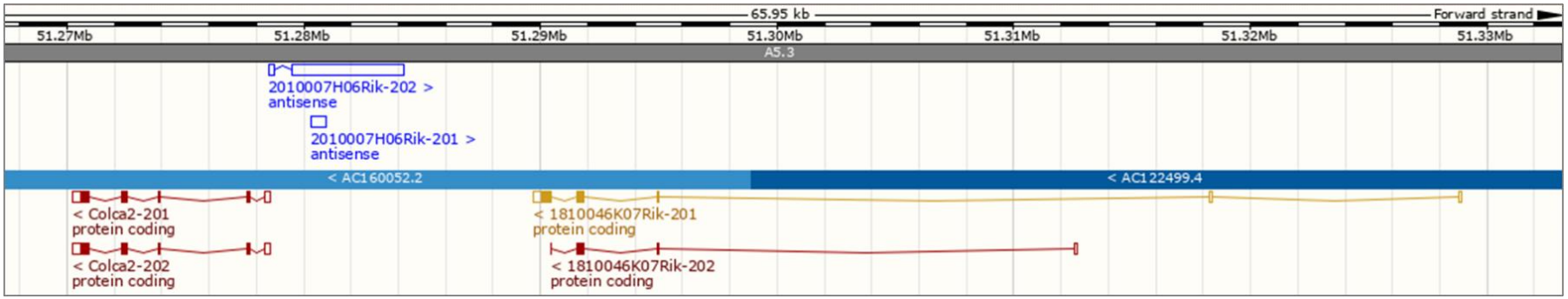


Figure 1.10: A schematic of the mouse locus 9A5.3, containing *18100046K07Rik* (orthologue of *C11orf53*), *2010007H06Rik* (orthologue of *C11orf92*/*COLCA1*), and *Colca2* (orthologue of *C11orf93*/*COLCA2*), showing their exon structure and relative positions and sizes. The position and orientation of the three genes to each other are conserved between human and mouse. Image from Ensembl, July 2018.

1.3a) *1810046K07Rik* transcripts							
Name	**Transcript ID**	**bp**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
1810046K07Rik-201	ENSMUST00000039959.10	1318	235aa	Protein coding	CCDS52792	Q9D8Q6	NM_027217 NP_081493
1810046K07Rik-202	ENSMUST00000181366.1	466	103aa	Protein coding	-	M0QWA8	-
1.3b) *2010007H06Rik* transcripts							
Name	**Transcript ID**	**bp**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
2010007H06Rik-202	ENSMUST00000186025.1	4934	No protein	Antisense	-	-	-
2010007H06Rik-201	ENSMUST00000050829.1	626	No protein	Antisense	-	-	-
1.3c) *Colca2* transcripts							
Name	**Transcript ID**	**bp**	**Protein**	**Biotype**	**CCDS**	**UniProt**	**RefSeq**
Colca2-202	ENSMUST00000215038.1	1269	239aa	Protein coding	CCDS57676	F8VPY8	-
Colca2-201	ENSMUST00000114427.3	1263	239aa	Protein coding	CCDS57676	F8VPY8	NM_001195681 NP_001182610

Table 1.3: Details of the known transcripts of mouse orthologues of (a) *C11orf53*, (b) *C11orf92* and (c) *C11orf93*. Data from Ensembl, July 2018.

1.3 Aims

The overarching aim of this project is to increase functional understanding of the 11q23.1 locus and its role in CRC, with a view to aid any future development of improved prevention and treatment of CRC in people with the risk genotype.

I will take two main approaches to this; the first, covered in chapters three and four, is to examine the localisation of the gene products. I will use immunological detection of the proteins in cell lines, via DDK-tagged overexpression proteins, to observe if the proteins are localising to specific subcellular compartments, in addition to working on generating a mouse line carrying endogenous fluorescent tags. I will also use in situ hybridisation to detect messenger RNA (mRNA) transcripts within human and mouse colonic tissue, to view the distribution of gene expression within the tissue context.

The second approach, covered in chapters five and six, is to assess the effect of loss of these genes on biological systems. I will generate and characterise a knockout mouse model of the locus, including histological assessment of tissue, differential expression analysis and tumour induction studies. I will also generate ex vivo cell and organoid models, using material from the knockout mouse, for further quantitative experiments on features such as cell proliferation.

Currently very little is known as to the function of *C11orf53*, *C11orf92* and *C11orf93*, and so we cannot know their role in CRC development beyond the observation that reduced expression is associated with an increase in risk. While the time limits of this project will necessarily preclude in depth characterisation of these genes, I aim to begin to narrow down the possible pathways involved to inform future, more targeted research into their activity.

Chapter 2: Materials and Methods

2.1 DNA Methods

2.1.1 DNA extraction from mouse earclips

Earclips were taken from mice at 14 days old by animal technicians and stored at -20°C until use.

To each was added 100µl DirectEar PCR lysis buffer (Viagen, 402-E) and 4µl 10mg/ml proteinase K solution (Promega, MC5005). Samples were incubated at 55°C for 14 hours and 85°C for 45 minutes. 300µl dH_2O was then added to each tube and samples were mixed by vortexing.

Samples were stored at -20°C.

2.1.2 Polymerase chain reaction (PCR)

DNA solutions extracted from earclips were then used in PCR reactions to genotype the animals. Primers were purchased from Sigma and IDT and were resuspended in dH_2O.

2.1.2.1 *C11orf* genotyping PCR

Reaction mix

5µl 2X DreamTaq PCR master mix (Thermo Fisher Scientific, K1081)

0.25µl 20µM forward primer

0.25µl 20µM reverse primer

2µl DNA solution

dH_2O to total volume 10µl

Thermocycler program

95°C for 4 minutes

32 cycles of: 95°C for 35 seconds

55°C for 40 seconds

72°C for 1 minute 5 seconds

72°C for 10 minutes

Cooled to 10°C at 0.1°C/second

PCR products were visualised under ultraviolet (UV) light on 1.5% agarose gel (Biogene, 300-300) containing 1:10000 ethidium bromide (Thermo Fisher Scientific, 15585011), run at 100V.

2.1.2.2 *Apc*$^{Min/+}$ genotyping PCR

Reaction mix

13µl 2X PCR master mix

3µl 20µM MAPC MT primer

3µl 20µM MAPC 15 primer

0.5µl 20µM MAPC 9 primer

3µl DNA solution

dH_2O to total volume 25µl

Thermocycler program

94°C for 4 minutes

31 cycles of: 94°C for 1 minute

54°C for 2 minutes

72°C for 3 minutes

72°C for 10 minutes

Cooled to 10°C at 0.1°C/second

PCR products were visualised under UV light on 2% agarose gel containing 1:10000 ethidium bromide, run at 100V.

2.1.2.3 Primer sequences

Name	Sequence
M53 CRISPR check F	TTAATGCTGACTGGGATACC
M53 CRISPR check R	GGAAATCCAGCTCAAAGATC
M92 CRISPR check F	AGATGCTCTATGTTAGTGCC
M92 CRISPR check R	CCTACCAAACAAAAGCCAAA
M93 CRISPR check F	ATGTGGCATTTAGATACTGG
M93 CRISPR check R	GGATCAAGGCTACGTATATG
M53 tag check F	TCAGTCCTACTCGCTGCATG
M53 tag check R	TGGCCTCCTTGTGAAGGTTG
M93 tag check F	GCTACTTCTGCCAACATTGG
M93 tag check R	CACAAGACCAAGTTTGAGGG
Shr2 tag check F	TGTTCGACATCCTGGCTACC
Shr2 tag check R	GTTGTTTAGCAAAGCCTGGC
MAPC MT	TGAGAAAGACAGAAGTTA
MAPC 15	TTCCACTTTGGCATAAGGC
MAPC 9	GCCATCCCTTCACGTTAG

2.1.3 Gel purification

Samples were run on agarose gel containing ethidium bromide and visualised using UV light. Bands were cut from the gel using a clean scalpel. DNA was purified from gel sections using QIAquick gel extraction kit (Qiagen, 28704), according to supplied kit instructions.

2.1.4 Restriction digest

Reaction mix

200µl 10X appropriate endonuclease buffer

0.75µl restriction endonuclease

0.2µl 100X bovine serum albumin (BSA) (New England Biolabs, B9000S)

1µg plasmid

dH_2O to total volume 20µl

Incubated at 37°C for 70 minutes in water bath and visualised on 1.5% agarose gel run at 60V, alongside 1µg uncut plasmid.

2.1.5 Annealing

Reaction mix

5µl each 100µM oligonucleotide

10µl dH_2O

Incubation

37°C for 30 minutes

95°C for 5 minutes

Cooled to 10°C at 0.1°C/second

Diluted 1:200

2.1.6 Ligation

Reaction mix

0.5µl cut vector

1µl diluted annealed oligonucleotide

2.5µl dH_2O

5µl 2X T4 ligase buffer (New England Biolabs, B0202S)

1µl T4 DNA ligase (New England Biolabs, M0202S)

30 minutes at room temperature (RT)

2.2 RNA Methods

2.2.1 RNA extraction from tissue

Mouse tissue was harvested, washed in phosphate-buffered saline (PBS) (IGMM Technical Services), and stored in RNAlater (Thermo Fisher Scientific, AM7021) at 4°C overnight. Tissue was then chopped up and moved to fresh tube. 1ml Trizol (Thermo Fisher Scientific, 15596026) and a 7mm ball bearing was added to each and the tissue homogenised in the TissueLyser using four 8 minute 50Hz cycles. Samples were left for 5 minutes at RT and moved to fresh tubes. 0.2ml chloroform was added and samples shaken for 15 seconds and left at RT for 10 minutes.

Samples were centrifuged at 12000g for 15 minutes at RT. The colourless upper aqueous phase was moved to a new tube, the white solid and red lower aqueous phases were discarded. 0.5ml isopropanol was added and left for 10 minutes at RT. Samples were centrifuged at 12000g for 10 minutes at 4°C and the supernatant discarded.

1ml 75% ethanol was added and samples were mixed by vortexing. Samples were centrifuged at 7500g for 5 minutes at 4°C, the supernatant discarded and the pellet air-dried. RNA was resuspended in 30µl dH_2O at 57°C, quantified using the Nanodrop, and stored at -80°C.

2.2.2 DNase treatment

RNA was DNase treated prior to cDNA synthesis. DNase reaction mixture was prepared on ice:

1µl DNase I (New England Biolabs, M0303S)

1µl 10X DNase buffer (New England Biolabs, B0303S)

1µg RNA

dH_2O to total volume 9µl

Samples were incubated at 37°C for 30 minutes.

1µl DNase STOP was added to each reaction and samples incubated at 65°C for 10 minutes.

2.2.3 cDNA synthesis

cDNA synthesis reaction mix was added to each 10µl DNase-treated sample:

4µl 5X M-MLV reverse transcriptase buffer (Invitrogen, 18057018)

2µl random primer mix (Invitrogen, 48190011)

2µl dNTP mix (10mM each) (Invitrogen, 18427088)

1µl RNase inhibitor (Thermo Fisher Scientific, N8080119)

1µl M-MLV reverse transcriptase (Invitrogen, 28025013)

Samples were incubated at 37°C for 1 hour, followed by 95°C for 5 minutes.

2.2.4 RNA synthesis

RNA was synthesised using HiScribe T7 RNA synthesis kit (New England Biolabs, E2050S) and 1µg of DNA template, according to supplied kit instructions.

2.2.5 RNAscope on FFPE tissue

2.2.5.1 RNAscope protocol

RNAscope was performed on 5µm formalin-fixed paraffin-embedded (FFPE) tissue sections on superfrost plus slides. Slides were prepared and pretreated for RNAscope 2.5 assay with RNAscope Pretreatment Reagents (ACDBio, 322000 and 322500), according to ACDBio user manual number 322452, with the following modifications: following boiling in Target Retrieval Reagents, rather than a brief wash in 100% ethanol, mouse tissue slides were left in 100% ethanol for 5 minutes and human tissue slides were left in 100% ethanol for 10 minutes (page 15); before addition of Protease Plus, slides were rehydrated by briefly dipping in PBS (page 15).

Singleplex RNAscope was performed using the RNAscope 2.5 HD Red Detection kit (ACDBio, 322360) and RNAscope Wash Buffer (ACDBio, 310091), according to ACDbio user manual 322360-USM, with the following modification: slides were stained with 50% Gill's

haematoxylin (Sigma-Aldrich, GHS132-1L) for 20 seconds, rather than two minutes (page 15).

Duplex RNAscope was performed using the RNAscope 2.5 HD Chromogenic Duplex Detection kit (ACDBio, 322500) and RNAscope Wash Buffer (ACDBio, 310091), according to ACDBio user manual 322500-USM, with the following modification: slides were stained with 50% Gill's haematoxylin for 20 seconds, rather than 30 seconds (page 20).

2.2.5.2 RNAscope probes

Gene	**Species**	**Channel**	**Catalogue number**
C11orf53	Human	C1	460841
C11orf53/1810046K07Rik	Mouse	C1	460861
C11orf92/COLCA1	Human	C1	447801
C11orf92/2010007H06Rik	Mouse	C1	460911
C11orf93/COLCA2	Human	C1	460851
C11orf93/Gm684	Mouse	C1	460871
Lgr5	Mouse	C2	312171-C2
PPIB (positive control)	Human	C1	313901
Ppib (positive control)	Mouse	C1	313911
Ppib (positive control)	Mouse	C2	313911-C2
DapB (negative control)	*E. coli*	C1	310043

2.3 CRISPR

2.3.1 Guide RNA preparation

Plasmid px458 was digested with restriction endonuclease BbsI (New England Biolabs, R0539S), as described above, using NEBuffer 2.1 (New England Biolabs, B7202S).

Guide RNA (gRNA) sequences were purchased as single stranded DNA oligonucleotides, together with reverse complements. Oligonucleotides were annealed, as described above. Sequences were designed so that annealed oligonucleotide would have sticky ends complementary to those left when plasmid px458 was linearised; a 5' G was added for T7 RNA synthesis, with an additional 5' G to reduce off-target effects.

gRNA oligonucleotides were ligated into linearised px458, as described above, and the plasmids transformed into competent cells, as described below. Colonies were picked and used in minipreps to obtain a stock of plasmid, as described below. Plasmids were Sanger sequenced by IGMM Technical Services to ensure correct insertion of the gRNA sequence.

gRNA template was then amplified from the px458 plasmids by PCR using T7 tagged forward primers and universal reverse primer. Product was gel purified, as described above, and subsequently used in re-PCR until the required concentration of 1µg in maximum 8µl.

T7-tagged templates were used for in vitro RNA synthesis using T7 RNA polymerase, as described above. gRNAs were purified using RNeasy kit (Qiagen, 74104), according to supplied instructions.

Name	**Sequence**
C11orf53 knockout gRNA sequence	TATGGAAGTCGGAATCGGAC
C11orf92 knockout gRNA sequence	GCAGTGGAAACCTGGCCTAA
C11orf93 knockout gRNA sequence	AGGACTCGAAGATCTGGTCG
C11orf53 mTFP1 tag gRNA sequence	TGGGGCTCTTATGAGTGC
C11orf93 EYFP tag gRNA sequence	GTTTCTACTAACTATAGC
Shroom2 mCherry tag gRNA sequence	GCATTAATCTCCAGGGCT
gRNA forward oligonucleotide	CACCGG[guide sequence here]
gRNA reverse oligonucleotide	AAAC[reverse complement guide sequence]CC
T7 tagged forward primer	TGTAATACGACTCACTATAGGG[guide sequence]
Universal reverse primer	AAAAGCACCGACTCGGTGCC

2.3.2 CRISPR injection mix preparation

RNA was all defrosted on ice.

Individual components spun at 13200rpm for 30 minutes at 4°C and the supernatant moved to a fresh tube.

Injection mix

50ng/µl Cas9 mRNA (Tebu Bio, L-6125-20)

25ng/µl each gRNA

150ng/µl each repair template

dH_2O to total volume 50µl

Mix was spun at 13200rpm for 30 minutes at 4°C and supernatant moved to a fresh tube.

Stored at -20°C and delivered to transgenic unit on ice.

2.4 Bacterial Methods

2.4.1 Transformation

Competent cells were defrosted on ice. 2.5µl ligation mix was added to each tube and mixed by swirling.

Cells were incubated

30 minutes on ice

42°C for 30 seconds using water bath

5 minutes on ice

950µl SOC medium (Thermo Fisher Scientific, 15544034) was added to each tubes and incubated on shaker in 37°C incubator for 1 hour.

Each culture was plated on two warm L-amp plates (IGMM Technical Services), one with 50µl and one with 200µl. Cultures were spread around whole plate using cell spreader. Plates were incubated at 37°C overnight and examined for colonies.

2.4.2 Plasmid purification

Colonies were picked into 5ml L-broth with ampicillin. Cultures were incubated on shaker in 37°C incubator overnight. Plasmids were extracted from expanded colonies using the QIAprep Spin Miniprep kit (Qiagen, 27104), according to supplied kit instructions.

2.4.3 Glycerol stocks

Colonies were picked into 5ml L-broth with ampicillin, prepared by adding 2µl ampicillin stock to 1ml LB. Cultures were incubated on shaker in 37°C incubator overnight. 0.85ml of 5ml overnight cultures were added to 0.15ml glycerol (Invitrogen, 15514-011). Tubes were vortexed and stored at -80°C.

2.5 Protein Methods

2.5.1 Protein extraction from cells

Lysis buffer

843µl RIPA buffer (Therma Fisher Scientific, 89900)

40µl complete EDTA-free protease inhibitor cocktail (Roche, 11873580001)

10µl pefablock (Roche, 11429868001)

1µl pepstatin (Sigma-Aldrich, P5318)

1µl Na_3VO_4 (Sigma-Aldrich, S6508-50G)

5µl PMSF (Thermo Fisher Scientific, 36978)

100µl phosSTOP (Roche, 4906845001)

Cell pellets were either freshly harvested and kept on ice, or defrosted on ice. Cells were lysed with 200µl lysis buffer per pellet on ice for 30 minutes, vortexing every 10 minutes. Lysates were centrifuged at 13200rpm for 30 minutes at 4°C and the supernatant moved to fresh tubes.

2.5.2 Protein quantification

Protein concentration was quantified by Bradford reagent (BioRad, 5000006) in 96 well plates. Standard curve was established by adding 0-7µl of 1mg/ml BSA (Vector Laboratories, SP-5050) were added to wells, in triplicate. Additionally 1µl of each sample was added to wells, in triplicate. 200µl Bradford reagent was added to each well and mixed by pipetting. The plate was read using an optical plate reader at 595nm. Sample concentrations were calculated from the BSA standard curve.

2.5.3 Western blots

2.5.3.1 Western blot solutions

4X resolving buffer

181.7g Tris base (Sigma, T6066)

20ml 20% sodium dodecyl sulphate (SDS) (IGMM Technical Services)

dH_2O to total volume 1000ml

pH 8.8

10% resolving gel

5ml 4X resolving buffer

4ml 40% acrylamide (Sigma, A7168)

9.8ml dH_2O

200µl 10% ammonium persulphate (APS) (Sigma-Aldrich, A3678)

20µl TEMED (Thermo Fisher Scientific, 17919)

4X stacking buffer

12.14g Tris base

4ml 20% SDS

dH_2O to total volume 200ml

pH 6.8

3% stacking gel

1.25ml 4X stacking buffer

0.375ml 40% acrylamide

3.375ml dH_2O

50µl 10% APS

5µl TEMED

10X running buffer

30g Tris base

144g glycine (Fisher Scientific, G/0800/60)

50ml 20% SDS

dH_2O to total volume 1000ml

50X wet transfer buffer

14.5g Tris base

72g glycine

dH_2O to total volume 500ml

1X wet transfer buffer

20ml 50X transfer buffer

780ml dH_2O

200ml 100% methanol

Loading buffer

1ml glycerol

500µl β-mercaptoethanol

2ml 10% SDS

1.25ml 1M Tris-HCl, pH 6.7

3ml 0.05% w/v bromophenol blue solution (Sigma-Aldrich, B0126)

2.5.3.2 Western blot protocol

7.2ml of 10% resolving gel added to 1.5mm mould, with 1ml dH_2O gently layered on top. Once gel had set, water was poured off, 3% stacking gel added to the top of the mould and comb inserted.

Protein sample

Required volume of protein

4µl loading buffer

dH_2O to total volume 20µl

Samples boiled at 100°C for 5 minutes in heating block and loaded into gel inside tank full of 1X running buffer. Gel was run at 100V for approximately 2.5 hours, until dye front had reach bottom of gel.

Polyvinylidene fluoride (PVDF) membrane (Amersham, 10600023) was soaked in 100% methanol until saturated. PVDF membrane, filter paper, sponges and gel soaked in 1X wet transfer buffer until saturated. Membrane and gel were layered inside filter paper and sponges in plastic holder and inserted into tank full of 1X wet transfer buffer. Transfer was run at 100V for 90 minutes at 4°C, with stir bar inside tank.

Membrane was washed in 1% PBS-tween (Sigma, P2287) (PBST) and blocked in 5% milk in 1% PBST for 1 hour at RT. Membrane was incubated with primary antibody in 3ml of 5% milk in 1% PBST overnight at 4°C on roller; antibody concentration is antibody dependent.

Membrane was washed 3 x 5 minutes in 1% PBST at RT. Membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibody diluted 1:2000 in 5% milk in 1% PBST for 1 hour at RT.

Membrane was washed 3 x 5 minutes in 1% PBST. Enhanced chemiluminescence (ECL) solutions 1 and 2 (Thermo Fisher Scientific, 32106) were mixed 1:1 and pipetted onto membrane, incubated for 1 minute at RT, and blotted off. Western blot was developed using light-reactive film (Amersham, 28906837).

To re-probe a membrane, membrane was washed 3 x 1 hour in 1% PBST before re-blocking.

HRP-conjugated anti-β-actin was used as loading control.

2.5.3.3 Western blot antibodies

Antibody	Manufacturer	Dilution
Anti-C11ORF53 (rabbit)	Sigma, SAB1103541	1:1000
Anti-C11ORF93 (goat)	Santa Cruz, Sc-245159	1:200
Anti-DDK (mouse)	OriGene, TA50011-100	1:1000
Anti-β-actin, HRP-conjugated	Santa Cruz, Sc-47778 hrp	1:50000
Anti-rabbit, HRP-conjugated	Promega, W4011	1:2000
Anti-goat, HRP-conjugated	Promega, V8051	1:2000
Anti-mouse, HRP-conjugated	Promega, W4021	1:2000

2.5.4 Immunofluorescence on cultured cells

2.5.4.1 Immunofluorescence protocol

Cells were grown on glass coverslips in six-well plates to 90% confluency. Cells were washed in PBs and fixed in 3ml per well of ice-cold 1:1 acetone:methanol at -20°C for 30 minutes. Fixed cells were washed 2 x 5 minutes in PBS on shaker at RT. 1ml 10% donkey serum (Sigma-Aldrich, D9663) in PBS was added per well and incubated for 1 hour at RT to block, followed by a brief PBS wash.

200µl of primary antibody solution, diluted according to individual antibody specifications in 10% donkey serum in PBS, was pipetted onto each coverslip and incubated for 1 hour at RT. Cells were washed 3 x 5 minutes in 1% PBST.

200µl of Alexa Fluor secondary antibody solution, diluted 1:1000 in 1.5% donkey serum in 1% PBST, was pipetted onto each coverslip and incubated for 20 minutes at RT in the dark. Cells were washed 2 x 5 minutes in 1% PBST.

Coverslips were mounted on superfrost plus slides using Vectashield mounting medium containing DAPI (Vector Laboratories, H-1200), and sealed around the edges with nail varnish. Slides were stored in the dark at 4°C until viewing using a fluorescence microscope.

2.5.4.2 Immunofluorescence antibodies

Antibody	Manufacturer	Dilution
Anti-C11ORF53 (rabbit)	Sigma, SAB1103541	1:300
Anti-C11ORF93 (goat)	Santa Cruz, Sc-245159	1:300
Anti-DDK (mouse)	Origene, TA50011-100	1:1000
Anti-calnexin (rabbit)	Cell Signalling, C5C9 #2679	1:150
Anti-calnexin (mouse)	BD Bioscience, 610524	1:400
Anti-γ-tubulin (mouse)	Sigma, T5326	1:1000
Anti-mouse Alexa Fluor 488 (goat)	Life Technologies	1:1000
Anti-rabbit Alexa Fluor 488 (goat)	Life Technologies	1:1000
Anti-goat Alexa Fluor 488 (donkey)	Life Technologies	1:1000
Anti-mouse Alexa Fluor 594 (goat)	Life Technologies	1:1000
Anti-rabbit Alexa Fluor 594 (goat)	Life Technologies	1:1000

2.5.5 Immunohistochemistry on FFPE tissue

2.5.5.1 Immunohistochemistry protocol

Immunohistochemistry was performed on sections of FFPE tissue by Vidya Rajasekaran, Calum Robertson and Marion Bacou.

Slides were rehydrated by incubation in xylene for 15 minutes and decreasing concentrations of ethanol (100%, 90%, 70%) for 10 minutes each at RT. Slides were dipped in dH_2O and blocked using 3% H_2O_2 solution for 20 minutes at RT. Slides were washed with 0.1% PBST for 5 minutes at RT.

Antigen retrieval was carried out suing 1X citrate buffer (Thermo Scientific, AP-9003-500) with 0.1% tween. Solution was heated to boiling point before slides were added. Slides were boiled in solution for 10 minutes. Slides were cooled at RT for 30-45 minutes and washed in 0.1% PBST for 5 minutes at RT. Slides were permeabilised with 0.5% Triton-X100 (Sigma-Aldrich, X100-500ML) in PBS for 20 minutes at RT and washed in 0.1% PBST for 2X 5 minutes at RT.

Slides were blocked with 5% donkey serum in PBST for 1 hour at RT. Primary antibodies were added to slides in blocking buffer and incubated for 2 hours at RT or overnight at 4°C. Slides

were washed in 0.1% PBST for 3X 5 minutes at RT. Secondary antibodies were added in blocking buffer and incubated for 2 hours at RT or overnight at 4°C. Slides were washed in 0.1% PBST for 5X 5 minutes at RT.

Signal was detected using by incubation with 3,3'-diaminobenzine (DAB) (Vector Laboratories, SK-4100) for 10 minutes at RT. Slides were washed in dH_2O for 5 minutes at RT and counterstained in Harris haematoxylin (Sigma-Aldrich, HHS32) for 30 seconds. Slides were washed under tap water for 5 minutes and dehydrated in increasing concentrations of ethanol (70%, 90% and 100%) for 15 minutes each and xylene for 20 minutes at RT. Coverslips were applied with mounting medium.

2.5.5.2 Immunohistochemistry antibodies

Antibody	**Manufacturer**	**Dilution**
Anti-Ki-67 (rabbit)	Abcam, ab15580	1:300
Anti-β-catenin (mouse)	BD Biosciences, 610153	1:100
Anti-cytokeratin-7 (rabbit)	Abcam, ab181598	1:500
Anti-Clca1 (rabbit)	Abcam, ab180851	1:300
Anti-rabbit, HRP-conjugated	Sigma, A9169	1:200
Anti-mouse, HRP-conjugated	Sigma, A9044	1:200

2.5.5.3 Other tissue staining

Haematoxylin and eosin (H&E) staining, Periodic Acid Schiff (PAS) staining and Alcian Blue staining was performed by the University of Edinburgh Division of Pathology.

2.6 Cell culture methods

2.6.1 Human cancer cell lines

2.6.1.1 Cell culture

Cells were cultured in DMEM (Gibco, 41965-039), supplemented with 10% fetal calf serum (FCS, IGMM Technical Services) and 1% penicillin-streptomycin (IGMM Technical Services). Cells were grown in T25 flasks in 5ml medium or T75 flasks in 12ml medium, with filter caps. Flasks were kept in incubators at 37°C and 5% CO_2. Cells were passaged at 90% confluency, 2-3 times per week.

To passage, medium was removed and cells were washed with warm PBS. Cells were detached from flasks by adding 2.0g/l warm trypsin solution (0.5ml for T25 flasks and 2ml for T75 flasks) and incubating at 37°C for 2-3 minutes. Following incubation, 2-4ml of warm DMEM was added to flasks and detachment of cells was ensured through pipette-washing and mechanical shock. Cells were pelleted by centrifugation at 3000rpm for 4 minutes. Cells were washed in warm PBS, re-pelleted, and resuspended in warm DMEM. Cell suspension was split into new flasks at 1:3.

If required, cells were counted using a Coulter counter. 100µl of cell suspension was diluted to 10ml in isotone. Particles were counted between 10µm and 30µm.

2.6.1.2 Transfection

Lipofectamine solution for one well of a six-well plate was made by adding 10µl of lipofectamine (Invitrogen, 11668-019) was added to 306.25µl Opti-MEM reduced serum medium (Gibco, 11058021), scaled up for required number of wells. Solution was incubated for 5 minutes at RT.

Plasmid solution for one well of a six well plate was made by adding 4µg of plasmid to 312.5µl Opti-MEM reduced serum medium, scaled up for required number of wells. Lipofectamine and plasmid solutions were mixed 1:1 and incubated for 30 minutes at RT.

Cells were washed with warm PBS and given 1.9ml DMEM supplemented with 10% FCS. 600µl of the transfection mix was gently added dropwise to each well. Non-transfected controls were given lipofectamine without plasmid.

2.6.2 Primary Mouse Fibroblasts

2.6.2.1 Fibroblast extraction

Complete medium was RPMI-1640 medium (Gibco, 21875-034) supplemented with:

10% FCS

1% penicillin-streptomycin

100µM asparagine (Sigma-Aldrich, A4159)

2mM glutamine (IGMM Technical Services)

50µM 2-mercaptoethanol

Pronase solution

10mg pronase protease (Merck Millipore, 53702)

494µl dH_2O

5µl 1M Tris pH8 (IGMM Technical Services)

1µl 0.5M ethylenediaminetetraacetic acid (EDTA) pH8 (IGMM Technical Services)

Solution incubated at 37°C for 30 minutes.

Collagenase-pronase solution

10mg collagenase D (Roche, 11088866001)

4ml complete medium

250µl pronase solution

Solution filtered through 0.22µm syringe filter.

1cm radius ear tissue was collected immediately post-mortem and stored in complete medium for transport from animal unit to tissue culture hood. Tissue was incubated in 70% ethanol at RT for 5 minutes and air-dried. Ear pieces were transferred to 10cm dishes containing 10ml of complete medium, and cut into pieces under 3mm radius. Tissue pieces were transferred into cryovials and collagenase-pronase solution was added to total volume of 1.8ml. Cryovials were incubated at 37°C for 90 minutes on a horizontal shaker at 200rpm.

Digested tissue was poured into 70µm cell strainers in 10cm dishes with 10ml fresh complete medium, and ground with syringe plungers for 5 minutes. The cell suspensions were moved into 50ml falcon tubes. Strainers and dishes were rinsed with further 10ml completed medium and this medium was also added to the falcon tubes.

Cell suspensions were centrifuged at 600g for 7 minutes at 4°C. The supernatant was discarded and the pellets resuspended in 10ml complete medium. Centrifugation and resuspension were repeated once. The cells were plated in 10cm dishes with 10µl amphotericin B (Sigma-Aldrich, A2492) at 250µg/ml. The following day cells were washed with PBS and given fresh complete medium with amphotericin B.

2.6.2.2 Fibroblast culture

When confluent, cells were washed with warm PBS and trypsinised by adding 2ml 1X EDTA-trypsin solution (Sigma-Aldrich, 59418C) and incubating at 37°C for 3 minutes. 6ml complete medium was added and cells suspensions centrifuged at 450g for 5 minutes at 4°C. Cell pellets were resuspended and plated at the desired density in complete medium. No amphotericin B was added after the first passage. When split 1:3 cells required passaging twice a week.

2.7 Cell Assays

2.7.1 SRB proliferation assay

Cells were seeded in 96-well plates in 200µl of medium per well, with one plate prepared per timepoint. Human SW480 colorectal cancer cells were seeded at 6000 cells per well and mouse primary fibroblasts were seeded at 1500 cells per well. In addition to seeded wells, wells of medium only were included to provide a blank. The outer ring of wells were filled with medium but were not included in calculations, instead providing an evaporation buffer.

At each timepoint the appropriate plate was mixed by adding 100µl 50% (w/v) trichloroacetic acid solution (Sigma-Aldrich, T0699) to each well and incubated at 4°C for 1 hour. The plate was rinsed four times with tap water and dried upside down on paper towels overnight.

When all plates were fixed 100µl 0.4% sulforhodamine B (SRB) solution (Sigma, S2902-100ml) was added to each well and incubated for 30 minutes at RT. Plates were washed with multiple brief 1% acetic acid washes until no stain was present in the wash. Plates were air dried overnight.

200µl 10mM Tris solution was added per well and the stain resuspended by either use of a plate shaker at 450rpm or manual pipette mixing. The plates were read on a plate reader at 565nm. The blank medium value was subtracted from the sample values.

2.7.2 Scratch migration assay

Cells were seeded in six-well plates in 10% FCS medium. When 90% confluent, the cells were washed with warm PBS and switched to 0.1% FCS medium. 24 hours later a scratch was made down the centre of each well with a P200 pipette tip and the cells washed and given fresh 0.1% FCS medium.

Brightfield images of the scratch were taken 1 hour, 24 hours and 48 hours post scratch. Marks drawn on the plate ensured the same region of the scratch was photographed at each timepoint. Closure of the scratch was measured using Selection and Measure tools in ImageJ to manually select the remaining gap and calculate it as a proportion of the original.

2.7.3 Chamber insert migration assay

Chamber inserts (ibidi, 80209) were adhered to the base of a 24 well plate. Cells were seeded at 49000 cells per chamber in 70µl of 10% FCS medium. At 90% confluence the cells were changed from 10% FCS medium to 0.1% FCS medium – cells were not washed with PBS due to the risk of dislodging the chambers. The following day the inserts were removed and the wells topped up with 300µl 0.1% FCS medium.

Plates were set up for live imaging using the Santana microscope, with one Brightfield image taken every 30 minutes for 72 hours. Closure of the gap was measured using Selection and Measure tools in ImageJ to manually select the remaining gap and calculate it as a proportion of the original.

2.8 Mouse methods

2.8.1 Tissue collection and processing

Mice were culled by cervical dislocation or CO_2 asphyxiation. Animals were pinned on their back and the body opened on the ventral midline. Organs were removed and rinsed in PBS before being placed in 10% neutral buffered formalin (NBF) (Sigma, HT501128). In the case of stomach, caecum and intestines, organs were flushed with PBS to remove all digestive material before fixation. Intestines were rolled onto 2mm diameter bamboo skewers, so that the outer serosa was against the skewer and the inner mucosa was exposed to the NBF. Tissue was fixed at RT for 24 hours, washed in PBS for 5 minutes at RT, and stored in 70% ethanol at 4°C. Tissue was then processed and paraffin embedded by the University of Edinburgh Division of Pathology.

2.8.2 Blood collection and processing

Animals were culled by CO_2 asphyxiation. Animals were pinned on their back and blood collected via cardiac puncture through the sternum, using a 30 gauge needle on a 1ml syringe. Blood was dispensed in 1.5ml tubes containing 10µl 0.5M EDTA anticoagulant per 300µl blood. Blood was mixed with EDTA by shaking. Tubes were stored on ice and delivered to the University of Edinburgh Easter Bush Pathology lab for automated blood counts.

2.8.3 1,2-Dimethylhydrazine injections

Mice were weighed the morning of the injection and their dosage calculated at 40mg 1,2-dimethylhydrazine (DMH)/kg body weight. DMH stock solution (Sigma, 40690) at 790mg/ml was diluted to 2mg/ml in 1mM EDTA-saline at pH 7.0. DMH was administered by subcutaneous injection.

Mice were weighed once a day to monitor for weight loss and checked twice a day for signs of being unwell, including hunched gait, reduced motility, rectal prolapse or blood around the anus. Animals were culled after one week and tissue harvested.

2.9 Organoid culture

Organoids were derived and cultured by Anna Maria Ochocka-Fox. Freshly harvested mouse intestines were flushed with cold PBS. Tissue was cut into 2-4mm pieces and washed with ice-cold PBS. Tissue fragments were incubated in either 3mM EDTA in PBS for the SI or 25mM EDTA in PBS for the LI, for 30 minutes at 4°C with gentle rocking. EDTA solution was removed and tissue resuspended in ice-cold PBS to release crypts. Tissue suspensions were passed through 70µm cell strainer to remove residual villous material. Tissue suspensions were then centrifuged at 300g for 5 minutes at 4°C to pellet crypts and allow removal of single cells. Pure crypts were added to warmed Matrigel (BD Bioscience, 356231) and plated in 24-well plates, with 2-4 small bubbles of Matrigel per well.

Organoids were cultured in advanced DMEM/F-12 medium (Gibco, 12634-028), supplemented with:

2mM GlutaMax (Invitrogen, 25050-079)

10mM HEPES (Invitrogen, 15630-056)

100µg/ml penicillin/streptomycin (IGMM Technical Services)

N2 supplement (Invitrogen, 17502-048)

B27 supplement (Invitrogen, 12587-010)

1mM N-acetylcysteine (Sigma-Aldrich, A9165)

0.1% BSA (Sigma-Aldrich, A9647-100G)

10nM [Leu-15] gastrin I (Sigma-Aldrich, G9145)

10nM nicotinamide (Sigma-Aldrich, N0636)

500nM A83-01 (Tocris Bioscience, 2939)

10µM SB202190 (Sigma-Aldrich, S7067)

50ng/ml murine recombinant EGF (Invitrogen, PMG8043)

100ng/ml murine recombinant noggin (Peprotech, 250-38)

1µg/ml murine recombinant R-spondin1 (R&D, 3474-RS-050)

100ng/ml muraine Wnt3A – LI only (R&D, 1324-WN-010)

Medium was changed twice weekly. Organoids were photographed on day one, two, four and eight of culture (with initial crypt isolation being day zero). Organoids were counted on day eight.

2.10 Statistics

Continuous data sets, such as RNAscope signal counts or intestinal polyp counts, were checked for normal distribution using the Shapiro-Wilks test. Comparisons between two data sets were then carried out using unpaired two-tailed student's t test, if both groups were normally distributed, or by two-tailed Mann-Whitney U test if not.

Categorial data sets, such as genotype distribution in breeding crosses, were compared using Chi-square test.

Chapter 3: Investigation of Gene Function through Protein Localisation

3.1 Introduction

The localisation of a protein both within individual cells and in the wider tissue context can give important hints as to function. There are many different cell types within the intestines, some of which are only found in certain regions of tissue, such as the stem cells being in the base of the crypt (Barker et al., 2007). By examining the expression patterns across tissue, and investigating possible co-expression with known cell-type markers, we can gain insight into the role of proteins of interest.

Further to establishing which cells express a protein, we can also look in detail within the cell to try and identify which subcellular component or organelle the protein of interest associates with, and if this localisation changes under different conditions, such as β-catenin's movement between the cytoplasm and nucleus in colonic cancers (Kobayashi et al., 2000).

The localisation of C11ORF92 and C11ORF93 has been previously examined in colonic tissue via immunohistochemistry (Peltekova et al., 2014). This study found that C11ORF92 occurs in granular organelles in immune cells of the lamina propria, including neutrophils, macrophages and mast cells, while C11ORF93 is present in the cytoplasm of cells from multiple origins, such as epithelial, lymphoid and mesenchymal. However, in silico analysis suggests that *C11orf92* is not protein coding (Smillie, 2015), which conflicts with an antibody-based approach to its localisation. Additionally, the anti-C11ORF93 antibody used in this study shows multiple bands on Western blot testing, although the authors speculate this is due to additional isoforms (Peltekova et al., 2014). They did not examine C11ORF53.

Western blot analysis by Claire Smillie on the available antibodies for C11ORF53, C11ORF92 and C11ORF93 showed that none have the high specificity desired for localisation studies (Smillie, 2015). As an alternative to direct detection of the proteins, I carried out immunofluorescence using overexpression plasmids carrying DDK-tagged versions of C11ORF53 and C11ORF93, and attempted to establish a mouse line with endogenous fluorescent tags. Due to the data indicating *C11orf92* does not produce a protein, it was not included in these experiments.

3.2 Use of antibodies to visualise protein localisation

3.2.1 Methods

3.2.1.1 Immunofluorescence

I performed immunofluorescence on proliferating human cancer cell lines, cultured on glass coverslips in six-well plates. These cells were stained with primary antibodies to proteins of interest, followed by fluorescent secondary antibodies. The cells were also stained with DAPI to mark the nuclei and mounted. Slides were imaged using fluorescence microscopy.

Claire Smillie has previously demonstrated that the available antibodies for C11ORF53 and C11ORF93 have issues of non-specific binding, visible as multiple extra bands on Western blots (Smillie, 2015). To confirm that the antibodies have the same lack of specificity when used for immunofluorescence as for Western blots, I used them to stain HCT116 and HT29 colorectal cancer (CRC) cell lines (figure 3.1). HT29 cells have high expression of all three *C11orf* genes, while HCT116 cells do not have any expression detectable by quantitative real time PCR (qRT-PCR) (Smillie, 2015). Both the anti-C11ORF53 and anti-C11ORF93 antibodies show signal across the HCT116 cells which, given the qRT-PCR data, we must consider to be non-specific. There is stronger signal in the HT29 cells, but when the images from the two cells lines are compared it is difficult to determine what signal in the HT29 cells is specific binding and what is background.

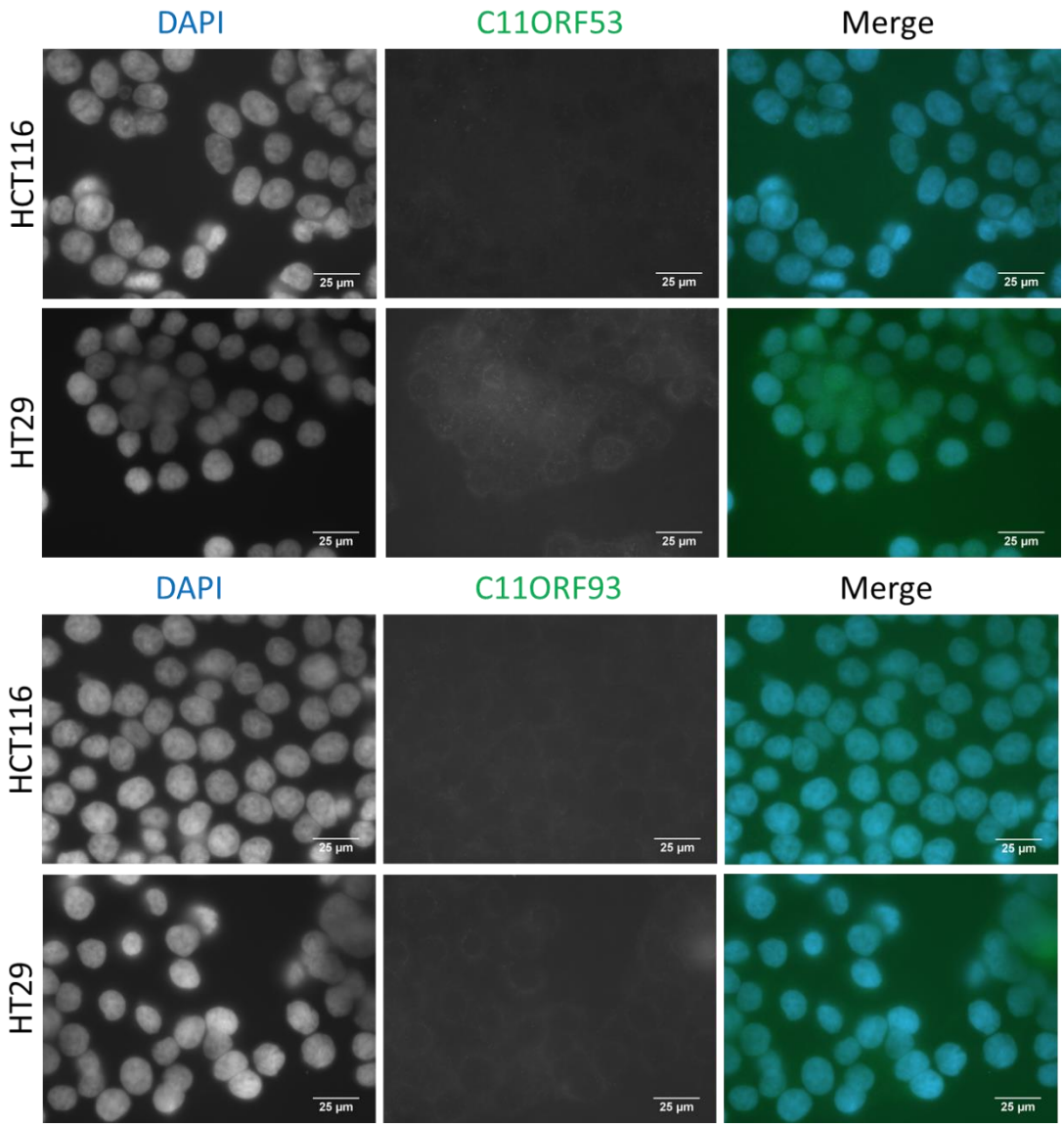


Figure 3.1: Fluorescence images of fixed proliferating HCT116 and HT29 colorectal cancer cells, stained with DAPI and either (a) anti-C11ORF53 or (b) anti-C11ORF93, using Alexa Fluor 488 secondary. HCT116 cells have no endogenous expression of either gene while HT29 cells have high levels of expression, as determined by qRT-PCR. As signal is visible in both cell lines, it therefore must be considered due to non-specific binding, and these antibodies are not suitable to visualise localisation.

3.2.1.2 Overexpression plasmids

To circumvent the lack of high specificity antibodies to the proteins, I utilised Myc-DDK-tagged overexpression plasmids. These plasmids, which were based on the pCMV6-entry vector, contained cDNA from either *C11orf53* or *C11orf93*, with a Myc-DDK tag on the 3' end (figure 3.2). The C11orf53 transcript used encodes the 236 amino acid protein. The *C11orf93* transcript used encodes the 154 amino acid protein; while there are other transcripts produced, at the time of purchase of these plasmids this was the only transcript available. Neither plasmid includes the 5' or 3' untranslated regions. The fusion genes had the CMV promoter, which gives strong constitutive expression in mammals, and the Kozak consensus sequence for transcription initiation. The plasmids also contained a neomycin resistance gene to allow selection. These plasmids were transfected into plated HeLa and SW480 cells; the empty pCMV6 plasmid was also used as a control.

Protein lysates from transfected cells were used in Western blot analysis for C11ORF53, C11ORF93 and DDK, which shows the transfected plasmids were successful in producing overexpressed DDK-tagged proteins (figure 3.3). While there are many extra bands detected by the anti-C11ORF53 and anti-C11ORF93 antibodies, it is still possible to see one band of increased intensity in the corresponding overexpression lysate, matching the DDK bands.

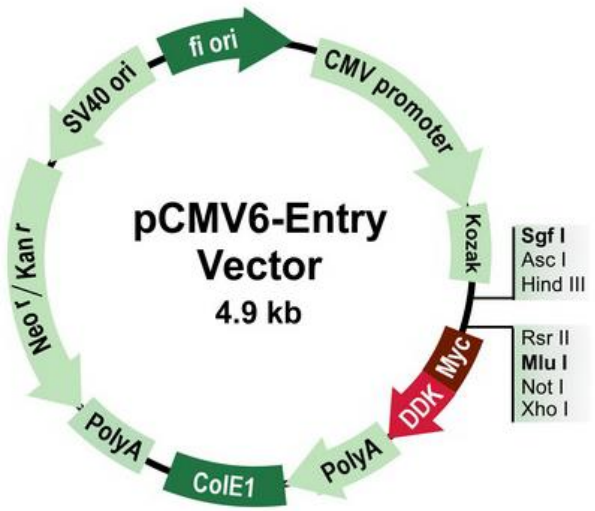


Figure 3.2: The pCMV6-Entry plasmid from OriGene transfected into cell lines for transient transfection. cDNA sequences of *C11orf53* transcript ENST00000637637.1 and *C11orf93* transcript ENST00000398035.6 were inserted between the Kozak sequence and the Myc-DDK tag.

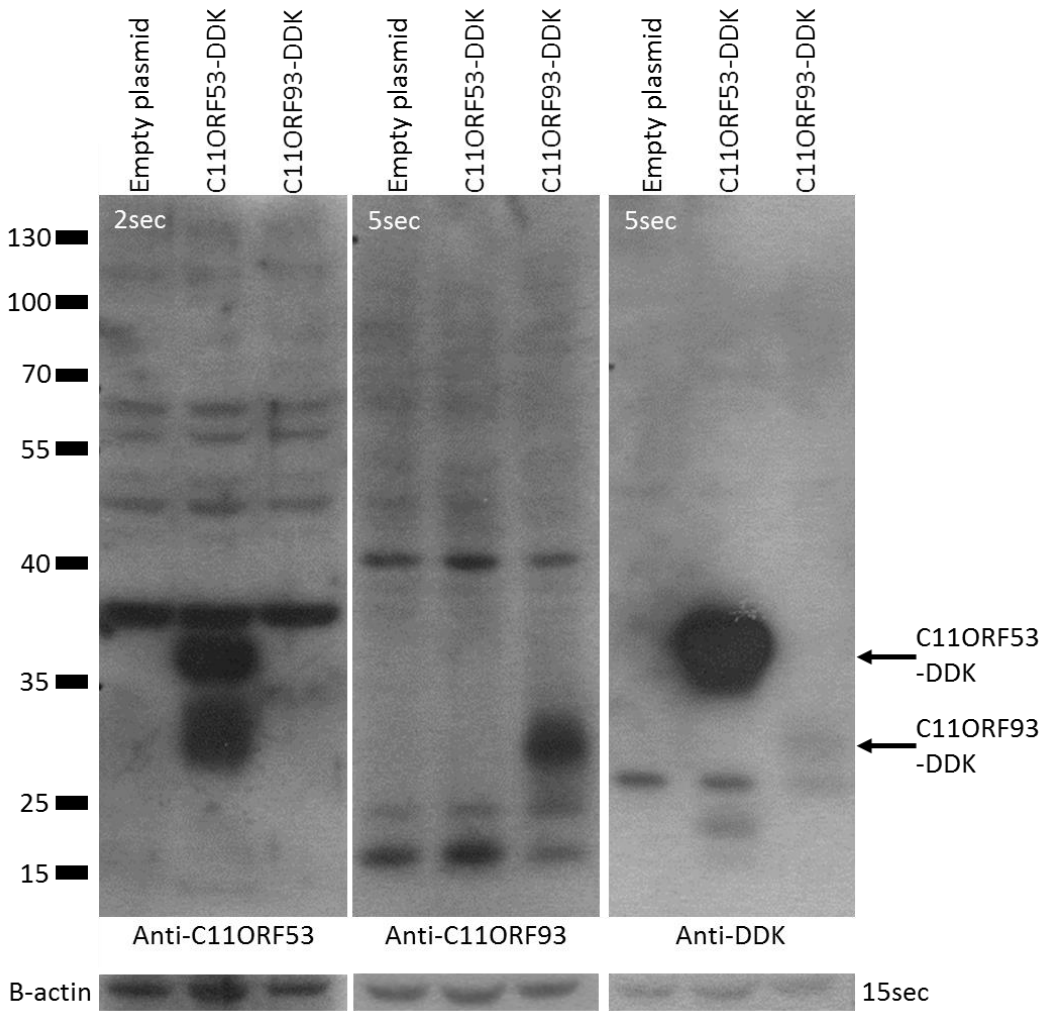


Figure 3.3: Western blots of SW480 cell lysates following transfection with either empty plasmid, C11ORF53-DDK overexpression plasmid or C11ORF93-DDK overexpression plasmid. Membranes were probed for C11ORF53, C11ORF93 and DDK. Predicted sizes of proteins are 29kDa for C11ORF53-DDK and 20.4kDa for C11ORF93-DDK, so the observed bands are bigger than expected; this was also observed by Claire Smillie. The C11ORF53-DDK plasmid gives higher overexpression than the C11ORF93-DDK plasmid, demonstrated by the different intensities of the DDK bands and the different exposure times required.

3.2.2 Results

3.2.2.1 Co-localisation of tagged proteins

I carried out immunofluorescence on SW480 and HeLa cells that had been transfected with the DDK-tagged overexpression plasmids. Cells were also stained with DAPI to identify the nuclei. SW480 is a CRC cell line with very little expression of the *C11orf* genes and is amenable to transfection. HeLa cells also have a low level of *C11orf* expression and are similarly receptive to transfection. This cell line is from a cervical cancer, rather than CRC, but was still included to compare if the proteins' activity in colonic cells is conserved in cells of other tissues. The transfection efficiency was not 100%, as evidenced by not all cells being positive for DDK. This efficiency could be improved by optimising the protocol, but as I was achieving sufficient overexpression to both see it on the Western blot and in enough of the cells to assess the localisation, it seemed unnecessary to spend time on optimisation.

As previous work suggested a link to the endoplasmic reticulum (ER) (Smillie, 2015), I co-stained for the ER protein calnexin alongside DDK. Calnexin is a molecular chaperone which retains misfolded or incompletely folded proteins in the ER (Bergeron et al., 1994). As it is an integral ER membrane protein, it can be used to define the area of the ER membrane. C11ORF53-DDK co-localises with calnexin in both HeLa and SW480 cells (figure 3.4a), implying C11ORF53-DDK is also located within the on the ER membrane or within the ER.

C11ORF93-DDK gives up to four punctate dots per cell of strong signal (figure 3.4b). These dots do appear within the bounds of the ER, as marked by calnexin, but as we are capturing 3D cells in 2D images we cannot say if the protein is within the ER or associated with the membrane, or simply above or below the ER. I also stained for γ-tubulin, which plays a role in microtubule nucleation (Oakley et al., 2015), as it forms similar foci, but there was no co-localisation between γ-tubulin and C11ORF93-DDK (figure 3.5).

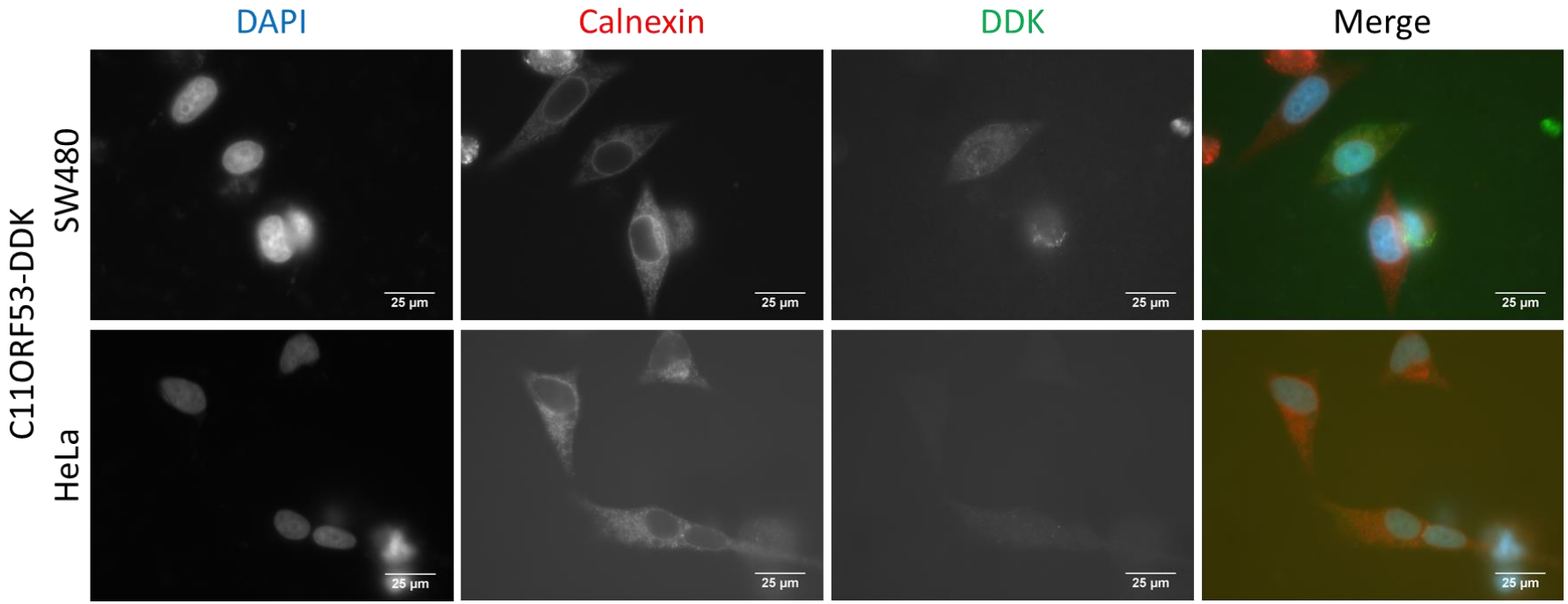


Figure 3.4a: Fluorescence images of proliferating SW480 and HeLa cells, transfected with pCMV6 plasmid containing C11ORF53-DDK. Cells were stained with DAPI, anti-calnexin using Alexa Fluor 594 secondary to mark the endoplasmic reticulum, and anti-DDK using Alexa Fluor 488 secondary to mark the transiently-overexpressed C11ORF53. The C11ORF53-DDK signal colocalises with calnexin, suggesting that C11ORF53-DDK is found throughout either the surface of or within the endoplasmic reticulum.

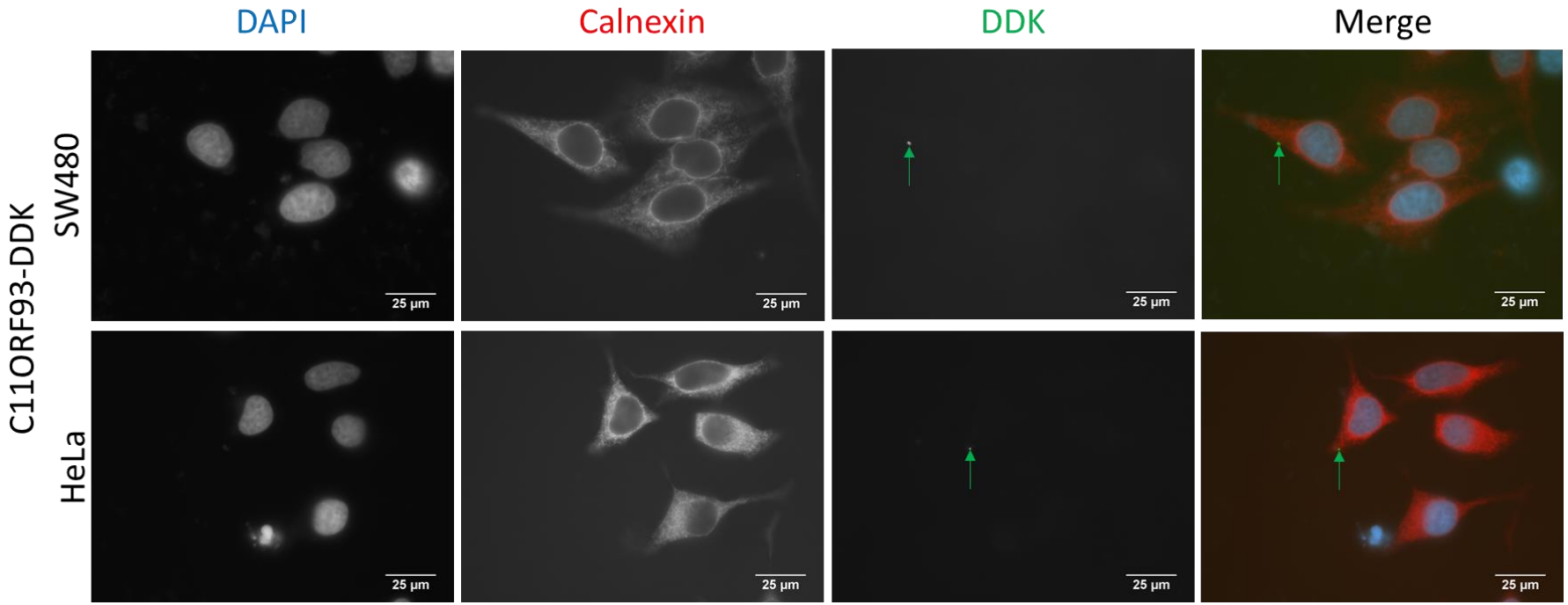


Figure 3.4b: Fluorescence images of proliferating SW480 and HeLa cells, transfected with pCMV6 plasmid containing C11ORF93-DDK. Cells were stained with DAPI, anti-calnexin using Alexa Fluor 594 secondary to mark the endoplasmic reticulum, and anti-DDK using Alexa Fluor 488 secondary to mark the transiently-overexpressed C11ORF93. The C11ORF93-DDK forms punctate dots, indicated with arrows, that appear within the endoplasmic reticulum as marked by calnexin, suggesting that these small regions of C11ORF93-DDK are found either on the surface of or within the endoplasmic reticulum.

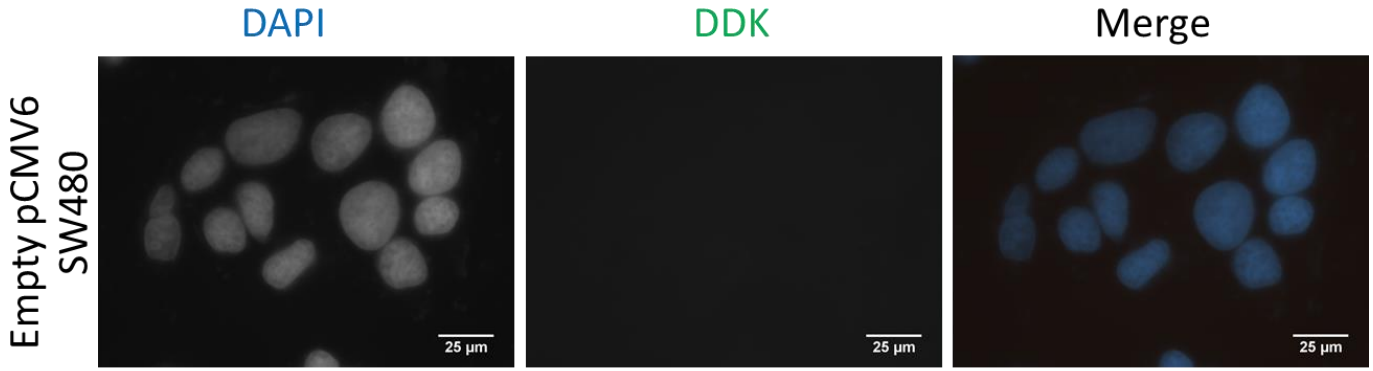


Figure 3.4c: Fluorescence images of proliferating SW480 cells, transfected with empty pCMV6 plasmid, as a control to the cells transfected with pCMV6 plasmid carrying *C11orf53* and *C11orf93*. Cells were stained with DAPI and anti-DDK using Alexa Fluor 488 secondary. The absence of DDK signal in these control cells shows that the signal observed in figures 3.4a and 3.4b is due to the behaviour of the desired tagged proteins.

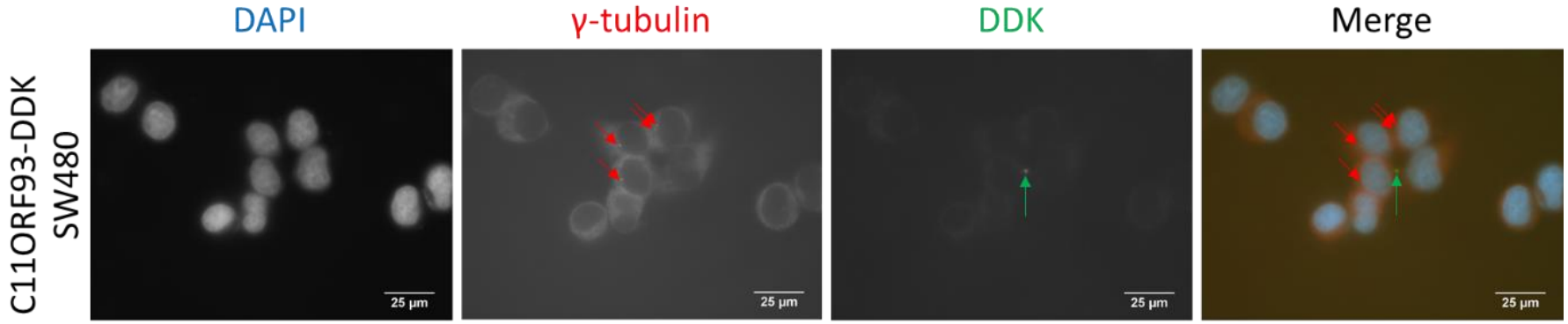


Figure 3.5: Fluorescence images of proliferating SW480 cells, transfected with pCMV6 plasmid containing C11ORF93-DDK. Cells were stained with DAPI, anti-γ-tubulin using Alexa Fluor 594 secondary and anti-DDK using Alexa Fluor 488 to mark the transiently-overexpressed C11ORF93. γ-tubulin and C11ORF93-DDK foci indicated with arrows. As the γ-tubulin and C11ORF93-DDK foci do not colocalise, C11ORF93-DDK is not associated with γ-tubulin within the cells.

3.2.2.2 Proliferation assay on transfected cells

In order to assess possible function of the proteins, I carried out a sulforhodamine B (SRB) proliferation assay on transfected cells to measure the effect of overexpression of the proteins. The assay was performed on SW480 cells that were either non-treated or had been transfected with the empty plasmid, C11ORF53-DDK overexpression plasmid or C11ORF93-DDK overexpression plasmid.

All cells transfected with a plasmid showed a substantially slower proliferation rate than non-treated cells (which can be expected, as transfection and overexpression can be highly stressful for cells), but there were no observable differences between transfection with the three different plasmids (figure 3.6). This is consistent with data from Claire Smillie showing no significant difference in proliferation when CRC cell lines are treated with siRNA to knock down C11ORF53 or C11ORF93 (Smillie, 2015), and my own data showing no significant difference in proliferation in primary mouse cell lines which have had the three genes knockout out by CRISPR (see chapter 6).

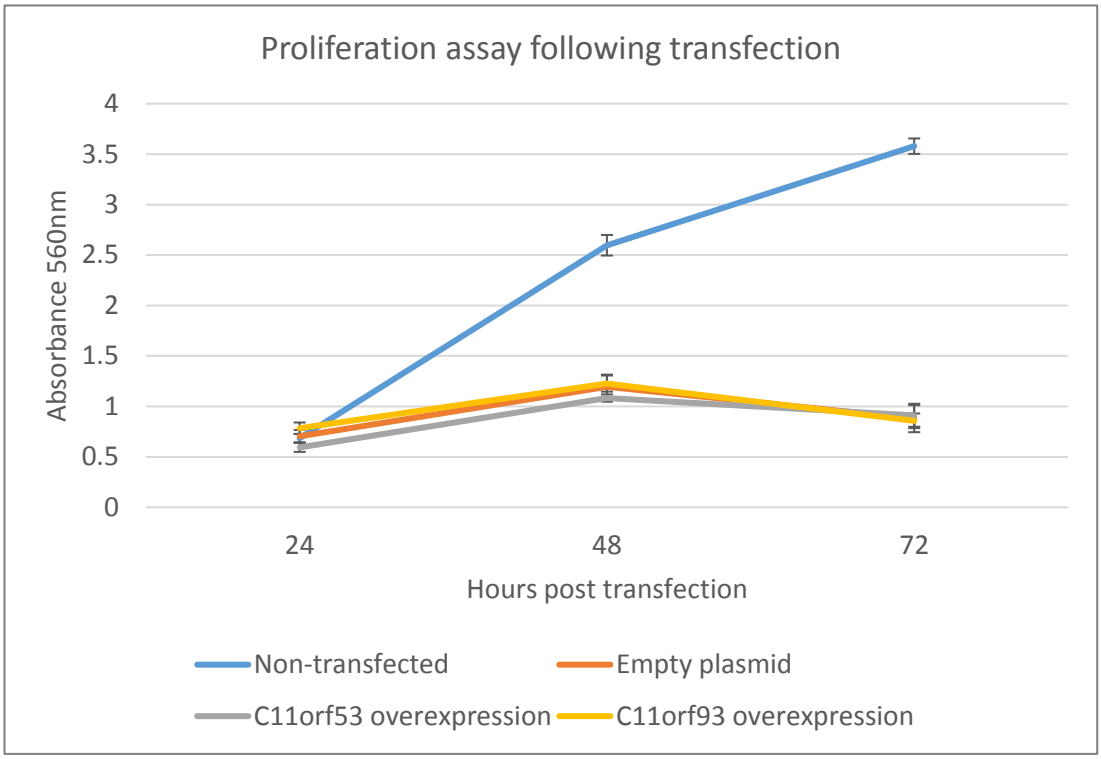


Figure 3.6: SRB proliferation assay on SW480 cells following transfection with either empty plasmid, C11ORF53-DDK overexpression plasmid or C11ORF93-DDK overexpression plasmid, with measurements taken 24, 48 and 72 hours post-transfection. Non-transfected cells also included for comparison. Results shown are the mean of three experiments, each containing 8 replicates per cell treatment.

3.3 Generation of endogenously-tagged mouse lines

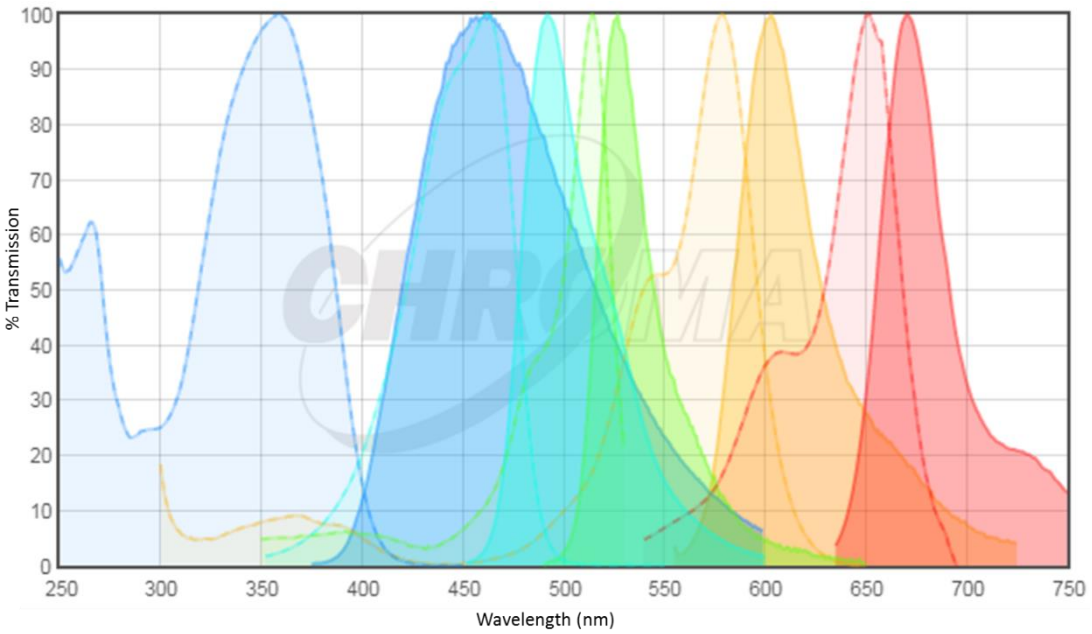
3.3.1 Methods

3.3.1.1 Choice of fluorescent tags

To remove the various caveats of using overexpression plasmids and obtain results we could have more confidence in, I undertook to create endogenously tagged proteins. This would keep the tagged protein under endogenous regulation, and the complete transcript, including untranslated regions (UTRs), would be produced. We chose to use fluorescent tags as it would remove the need for antibodies for visualisation, although fluorescent tags are larger than some other tags such as DDK.

Two different fluorescent proteins were chosen, to enable us to visualise them in the same sample; as we do not know how, or if, the proteins interact, this would allow observation of any patterns of co-expression in the same cell type or co-localisation to the same subcellular component. Following discussion with Ann Wheeler from the IGMM Advances Imaging Resource, the tags used were monomeric teal fluorescent protein 1 (mTFP1) for C11orf53 and enhanced yellow fluorescent protein (EYFP) for C11orf93. The wavelengths of these two fluorescent proteins allows co-visualisation of the tagged proteins with DAPI to mark the nuclei, additional immunofluorescence using an Alexa Fluor 568 secondary antibody, and the RNAscope in situ hybridisation technique using Cyanine5 (figure 3.7).

Rather than join the tags directly onto the protein, I included a short flexible linker peptide, $(GGGGS)_3$. By allowing the two proteins some distance and independent movement, the linker improves folding and stability and so aids function of both the protein of interest and the tag (Chen et al., 2013a).



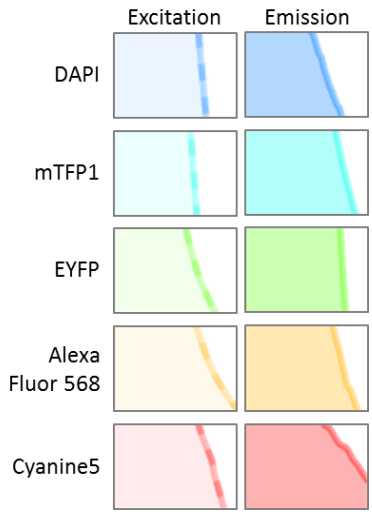


Figure 3.7: Excitation and emission spectra of the proposed fluorescent protein tags for C11ORF53 and C11ORF93, mTFP1 and EYFP. Use of these two proteins allows simultaneous visualisation of both tags alongside DAPI for nuclear staining, Alexa Fluor 568 fluorophore for immunofluorescence, and Cyanine5 fluorophore for RNAscope. Key to excitation and emission graph colours for each fluorochrome is shown on the left. Graph was generated using Spectra Viewer by Chroma, https://www.chroma.com/spectra-viewer.

3.3.1.2 CRISPR experiment design

While adding fluorescent tags could have been performed in cell lines, we are interested in localisation within the tissue context as well as the cell context. We therefore began by creating a tagged mouse line, as we could subsequently generate primary cell lines from the mice if required.

Due to the close proximity of *C11orf53* and *C11orf93*, it would be unlikely that we would obtain both tags on the same chromosome copy from breeding together tags on different chromosome copies. Breeding could still produce mice carrying both tags heterozygously, but the expression of these genes is already low and I was concerned any co-expression/co-localisation studies of the two proteins would be hampered if we could further only visualise half of the expressed protein. We therefore decided to target both genes in a single experiment to generate a compound tagged line.

It has been shown that CRISPR/Cas9 can be used to insert a tag immediately before the stop codon of a protein (Li et al., 2014); I used this approach and prepared guide RNAs (gRNAs) and repair template DNA to inject with Cas9 mRNA into mouse embryos. The guides were designed to target Cas9 cleavage as close to the stop codon as was possible with the available protospacer adjacent motif (PAM) sites. An additional 5' G was added to the sequence to reduce off-target effects (Cho et al., 2014).

The repair templates consisted of the linker peptide and fluorescent tag sequences flanked by 150bp of sequence either side of the stop codon (figure 3.8a, 3.8b). This 150bp length was recommended for short targeted changes, such as changing a SNP from one allele to another; I did consider if the homology arms should increase in length if a larger sequence is being inserted, but there was no clear guidance in the literature and, as the tags are all around 700bp, I was already concerned about the size of the repair template to be injected into the cells. The PAM sequence in the repair template was mutated to avoid Cas9 cleaving the template or re-cleaving the edited gene.

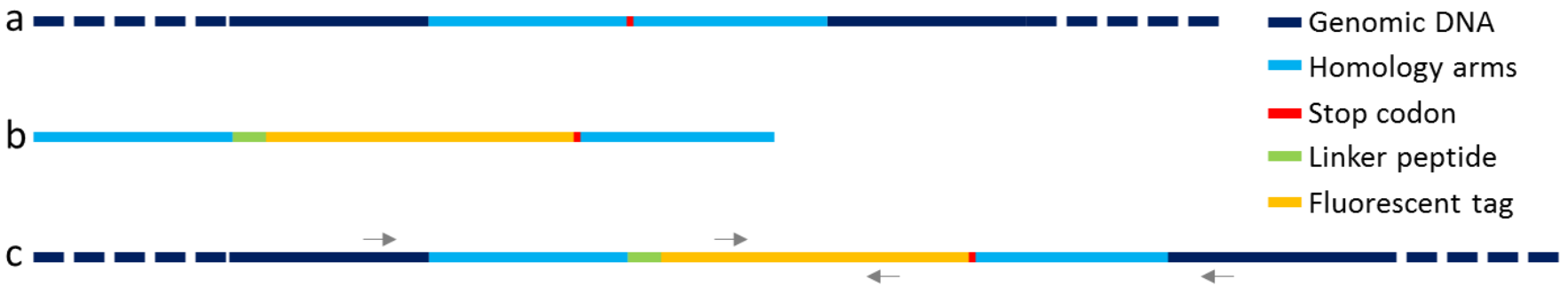


Figure 3.8: Schematic of CRISPR approach to inserting fluorescent tags at the C-terminus of C11ORF53 and C11ORF93. (a) The unedited gene sequence within the chromosome. The Cas9 endonuclease was targeted to within 10bp of the stop codon. (b) The repair template injected into the nuclei of C57BL6/J mouse eggs by microinjection. The repair template consisted of the final 150bp of gene sequence before the stop codon, the linker peptide sequence, the fluorescent tag sequence, the stop codon, and the first 150bp of sequence following the stop codon. (c) The edited gene resulting from homologous recombination following Cas9 activity. The linker peptide and fluorescent tag have been inserted immediately before the stop codon, leaving all other gene sequence unchanged. Genotyping primers, positions marked with grey arrows (not to scale), were used to determine if integration of the repair template had successfully taken place. The inner pair of primers were used to confirm the new sequence was present in the genome by the presence or absence of a PCR product, while the outer pair of primers were used to confirm the insertion was in the desired location by the size of the PCR product.

.3.3.1.3 Genotyping of potentially tagged animals

To identify animals with the fluorescent tag inserted, I designed primers within the tag sequences and 50bp either side of the homology arms (figure 3.8c). The inner primers validate that the fluorescent protein has integrated into the genome, as these sequences are not otherwise found in the mouse genome. The outer primers allow confirmation that the sequence has inserted in the correct place: the wild type product is approximately 400bp, but the presence of the tag gives and additional 700bp. Following amplification, sequencing can also ensure the insertion is in frame, so that the C-terminus of the protein and the fluorescent tag would be properly translated.

3.3.2 Results

3.3.2.1 Tagging C11orf53 and C11orf93

The CRISPR mix was injected into the nuclei of C57BL/6J eggs by Jacek Mendrychowski from the CBS Transgenic Core. These eggs were fertilised and kept overnight in an incubator; embryos at the two-cell stage were then transferred into females. We had three injection sessions, as the first two did not produce pregnancies.

Given this lack of pups, following the third session some embryos, both injected and non-injected, were kept in the incubator rather than being transferred into a female, so that Jacek Mendrychowski could monitor their development. All non-injected embryos progressed correctly to the blastocyst stage, but none of the injected embryos did. Approximately half of them died, although they did not lyse, while the other half appeared alive and healthy but were stalled at the two-cell stage. However, this third session did result in a successful pregnancy that gave two pups.

When the animals were genotyped, I saw only the 400bp wild type product using the primers outside the homology arms instead of the 1100bp product indicating the presence of the insertion. The primers within the tags gave no products. The additional check of mixing the primer pairs – amplifying from upstream of the 5' homology arm to within the tag, and from within the tag to downstream of the 3' homology arm – also gave no products.

From these PCR results we concluded the experiment had not worked and neither of the mice carried endogenous fluorescent tags.

3.3.2.2 Tagging Shroom2

Another gene of interest within the research group is *SHROOM2*, a gene on the X chromosome which has also been linked to CRC risk. As with C11orf53 and C11orf93, the antibodies available for SHROOM2 have issues of non-specificity that limit our ability to visualise the protein. I therefore also designed a CRISPR/Cas9 experiment to add an endogenous mCherry tag to Shroom2 in mice, and prepared the Shroom2-mCherry injection mix in parallel to the C11orf-tag injection mix. I used the same approach in designing primers for genotyping.

In contrast to the C11orf-tag experiment, 25 pups were born following the initial Shroom2-mCherry injection session. Genotyping suggested that five of these had mCherry inserted in the correct place and subsequent organoid culture shows the presence of red fluorescent signal (figure 3.9).

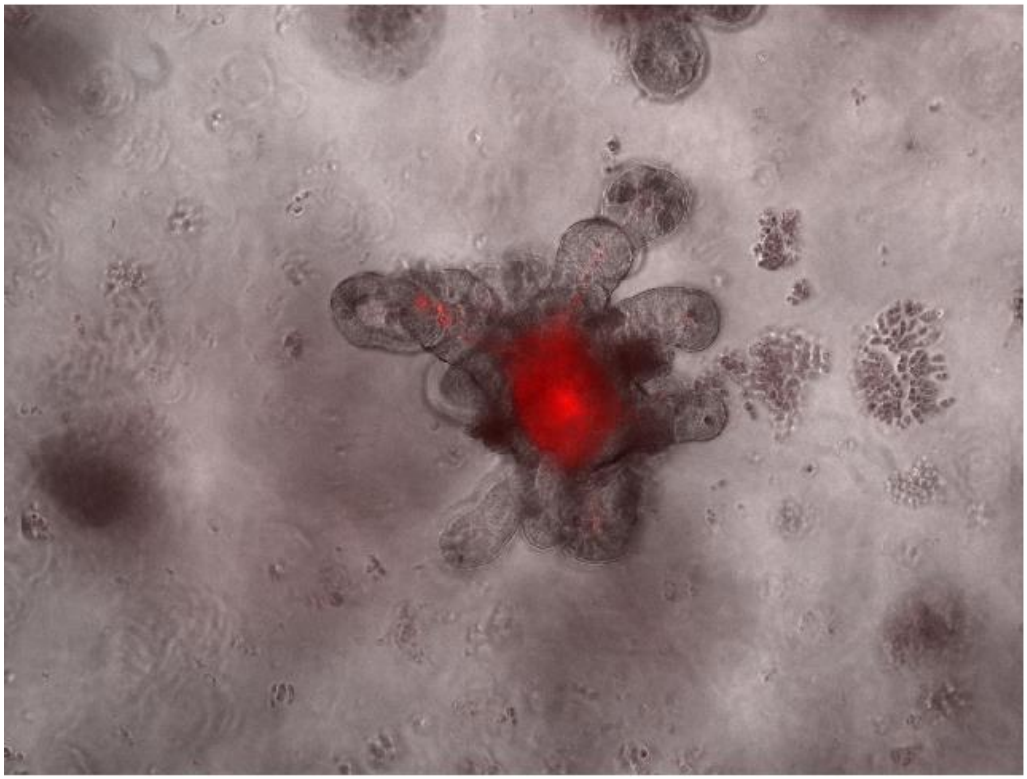


Figure 3.9: Merged fluorescence and brightfield image of an organoid derived from the small intestine of a SHROOM2-mCherry mouse, showing red fluorescent signal. Image taken at 20X. Image from Anna Maria Ochocka-Fox.

3.4 Discussion

To understand the function of *C11orf53* and *C11orf93*, I examined the localisation of their proteins. Due to the absence of high-specificity antibodies to either protein to detect them directly in cells or tissue, I took alternative approaches to visualise them, using protein tags.

C11ORF53-DDK co-localises with calnexin, which is consistent with previous data suggesting a connection to the ER. There are also many links between the ER and CRC, with ER stress affecting intestinal stemness, tumourigenesis, and inflammatory bowel disease (Drake et al., 2015; Heijmans et al., 2013; Hosomi et al., 2015; Niederreiter et al., 2013).

The nucleus is clearly defined within the calnexin staining as a region of no signal, but this is not the case for the DDK staining – there is little distinction between the nucleus and the surrounding ER. It is difficult to say if this means C11ORF53-DDK functions differently within the ER than the membrane protein calnexin, or that C11ORF53-DDK is present within the nucleus as well as the ER. It may also be peculiarity of the staining and bleedthrough with DAPI, although as there are no cells where the nucleus is stained and the ER is not, and the nuclei are not highlighted in the C11ORF93-DDK cells, I do not believe this is the case.

C11ORF93-DDK forms distinct foci, which appear within the bounds of the ER on 2D imaging but are not necessarily located within the ER in the 3D cell. It has been suggested that the foci resemble γ-tubulin, which is associated with centrioles and the basal bodies that anchor cilia, but subsequent co-staining shows the proteins do not co-localise. It would be interesting to do more exploration into what these foci are by investigating other potential co-localising proteins, but I chose to prioritise developing and characterising the mouse model, results from which could then be used to guide such study in a more targeted manner.

As these immunofluorescence results were obtained using overexpression vectors, there are several caveats and they should not be accepted unconditionally. The proteins are not being expressed at endogenous levels or under endogenous control; the ORF does not include the UTRs, and the C11ORF93 plasmid does not contain the complete cDNA sequence of the longer two transcripts; there is a Myc-DDK tag on the C-terminus; and the cells are under stress from the transfection and forced overexpression. Additionally, as the tag is not fluorescent, I had to use antibodies to visualise it. While the anti-DDK antibody looks highly specific on the Western blots, it is possible some non-specific binding could occur under immunofluorescence conditions.

It is encouraging that I am not seeing the same pattern with C11ORF53-DDK and C11ORF93-DDK, as it suggests the results seen are not purely an artefact of the plasmid, tag or antibody. However, as the plasmids carry tagged versions of two different proteins they may not show the same artefacts. The co-localisation of C11ORF53 and calnexin is supported by previous data from Claire Smillie suggesting a link between the *C11orf* genes and the ER, and while this does not rule out that the observed activity is being influenced by the artificial nature of the system, it does give supporting evidence that we can have confidence in it.

Given the limitations of an artificial overexpression system, endogenous tags would be highly useful. It is unfortunate that so far we have not successfully produced a mouse line with fluorescently tagged proteins, due to the very low survival of embryos that received the injection of the CRISPR mix.

We did not carry out any experiments to determine the cause of the developmental block that was observed in the injected embryos. However, the Shroom2-mCherry mouse line was successfully generated in tandem to the failed C11orf-tag lines. We can exclude the physical microinjection, due to the survival of both the Shroom2-mCherry line and other procedures carried out by the transgenic service during this period. Within the injection mixes from the two experiments, the Cas9 mRNA was aliquoted from the same batch, the gRNAs were purchased at the same time, and the repair templates were prepared in parallel. We can therefore assume the chemical composition between the two experiments is comparable, with the exception that as two proteins were being targeted in the C11orf-tag experiment there was a higher concentration of nucleic acids than in the Shroom2 experiment.

It is therefore likely that the developmental block observed in the C11orf-tag embryos is a genetic effect caused by the action of Cas9, rather than a chemical effect of the injection mix. The aim of this experiment was not to alter the function of the genes, but simply to add C-terminal fluorescent tags. It is possible that there were more substantial effects on *C11orf53* and *C11orf93* than intended. However, we can compare these results to the use of CRISPR to create a knockout mouse line (see chapter 5). Such massive developmental defects have not been observed in that complete knockout line, either during its creation or in subsequent generations, which suggests that in addition to any disruption that may have occurred to the targeted genes, the gRNAs used here have significant off-target effects.

As localisation of proteins can give key insights into their function, continued study to obtain more data into C11ORF53 and C11ORF93 localisation would be of great use in understanding

their role in CRC. Further use of the DDK-tagged plasmids, a new CRISPR experiment to add fluorescent tags using redesigned gRNAs, or trying to raise different antibodies to the proteins are all potential avenues for future work.

Chapter 4: Investigation of Gene Function through mRNA Localisation

4.1 Introduction

To investigate the function of the *C11orf* genes, and so hopefully elucidate some aspect of their cellular role, I examined the localisation of their mRNA within the intestine. For *C11orf92*, with no known protein, this is the only gene product that can be studied. *C11orf53* and *C11orf93* do produce proteins, but with all tested antibodies showing too much non-specific binding to be used in localisation studies, and no success with endogenous tags, mRNA was the only option available for viewing patterns of gene activity within colonic tissue.

mRNA is not interchangeable with protein, as levels of one do not necessarily reflect levels of the other, and the location of a protein itself within the individual cells can be significant. However, mRNA localisation is both a useful tool in its own right and a particularly valuable source of data when protein-based information is difficult to obtain. Within a complex organ such as the intestine, there are many cell types of different function, and these can be identified by their transcription profile (Grün et al., 2015). Any information we can gather into which cell types express our genes of interest can aid our understanding of how alteration in their expression can lead to CRC.

The expression patterns of the three genes were assessed using RNAscope on formalin-fixed paraffin-embedded (FFPE) colonic tissue from human CRC patients, wild type (WT) mice and $Apc^{Min/+}$ mice. The $Apc^{Min/+}$ mouse line (*Adenomatous polyposis coli*$^{Multiple\ intestinal\ neoplasia/+}$) carries a mutation which makes them highly susceptible to the spontaneous formation of intestinal adenomas, and is used as a model of human familial adenomatous polyposis (FAP) (Su et al., 1992). The mouse tissue was freshly harvested and fixed in 10% neutral buffered formalin (NBF) by myself and Vidya Rajasekaran, and processed and paraffin-embedded by the University of Edinburgh Division of Pathology at the Western General Hospital. Our Swiss roll technique of intestinal tissue preparation allows the entirety of both small and large intestine from a mouse to be treated and viewed on a single slide, so I was able to collect data along the whole length of the intestine. Human tissue was obtained from ten patients

who were undergoing surgery at Western General Hospital and consented to their tissue being used in research.

4.2 Utilisation of RNAscope to detect transcripts

4.2.1 Methods

RNAscope is an in situ hybridisation technique to visualise mRNA molecules, utilising a "double Z" probe pair design (figure 4.1). Both probes within a pair must bind exactly adjacent to provide the foundation for subsequent amplification, giving the technique high specificity. A pool of approximately 20 probe pairs is designed across the transcript and a signal can be detected as long as any three pairs correctly bind, giving high sensitivity even if some regions of the sequence have degraded or are inaccessible due to bound proteins. Following the multiple amplification steps, a single mRNA molecule can be visualised as a dot using a light microscope.

As the technique can be applied to FFPE tissue, the mRNA can be visualised and quantified within the tissue context. This offers a huge advantage over other quantitative techniques such as qRT-PCR analysis, which require destroying the tissue to extract the mRNA. In addition to expression patterns within regions of tissue, it is possible to accurately distinguish tissue types and quantify the expression within each individually, such as between crypt epithelium, lymphoid tissue, and underlying muscularis (figure 4.2). By comparison, in mRNA extraction techniques, while an effort can be made to only process stripped mucosa, it is far more difficult to be confident in having a pure cell population.

RNAscope was carried out on FFPE sections of human, WT mouse and *Apc*$^{Min/+}$ mouse tissue. Following the in situ hybridisation, slides were scanned using the NanoZoomer slide scanner from Hamamatsu, which allows the entire slide to be captured in one large high-resolution file, and the tissue examined. While there are options for automating dot counting, such as by Definiens or Image J, we did not have much success applying these to our images as we had great difficulty setting the parameters to achieve the correct sensitivity. This was further complicated by the way that probe pools for different genes can give dots of subtly different appearance; for example, *C11orf92* dots tend to be larger and have less defined edges than *C11orf93* dots, which are quite small and sharp. Additionally, the NanoZoomer obtains its large images as a series of smaller tiles, and the extent to which each of these is in focus can vary across a slide.

I therefore carried out manual counting on all of my slides. While I will not be 100% accurate in my signal identification, by having a clear definition of what I consider a dot I was able to

have internal consistency. This was verified by taking a subset of human patient normal mucosa and recounting all of their crypts for each gene several months after the initial count (figure 4.3). I was also able to be very precise in my delineation of the tissue regions of interest, to only count the desired cell types.

Across the tissues – human, WT mouse and *Apc*$^{Min/+}$ mouse – I counted the amount of signal for each *C11orf* gene in the intestinal crypts. I only counted properly sectioned crypts, meaning those where the whole length of the crypt was clearly visible (figure 4.4). Crypts were selected at a magnification where I could determine if a crypt was sectioned correctly but not visualise any RNAscope signal, to avoid bias in my crypt selection, before then using a higher magnification to perform the counts. To be able to quantify the distribution of transcripts within crypts, I split each crypt into thirds. This was not based on any biological markers, such as the presence of *Lgr5*$^{+}$ cells, but was a purely spatial division.

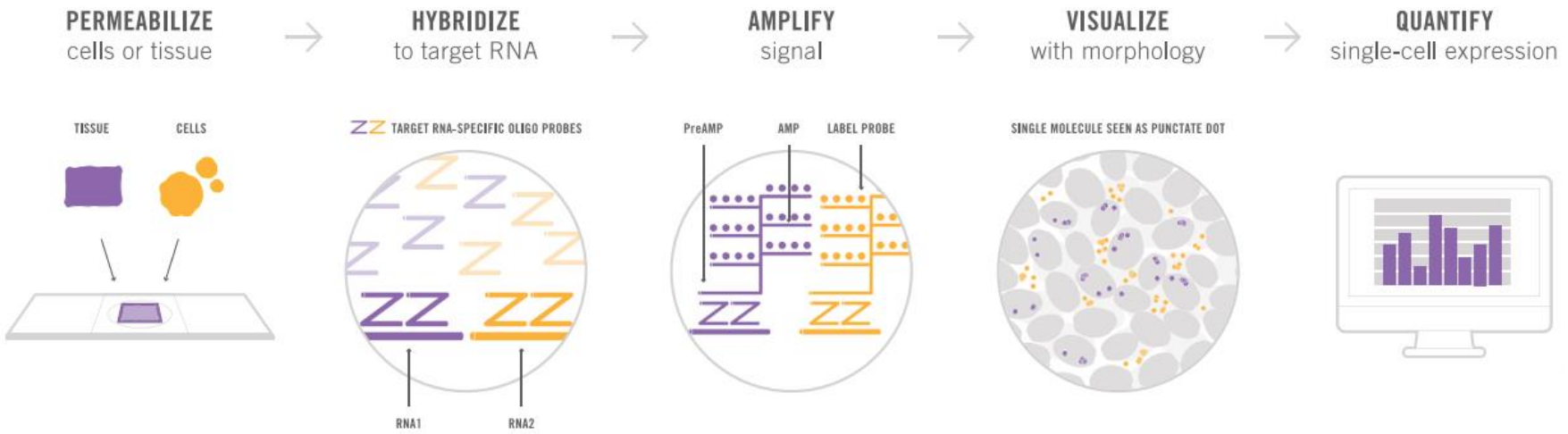


Figure 4.1: Schematic of RNAscope workflow. Fixed cells or tissue sections are permeabilised and probes are hybridised to mRNA. Both probes in the "double Z" pair must bind exactly next to each other to provide the base for amplification. A series of amplification steps allows a high amount of label to be bound to each mRNA molecule, and so individual mRNA molecules can be visualised as punctate dots within the sample. Diagram taken from ACDBio, https://acdbio.com.

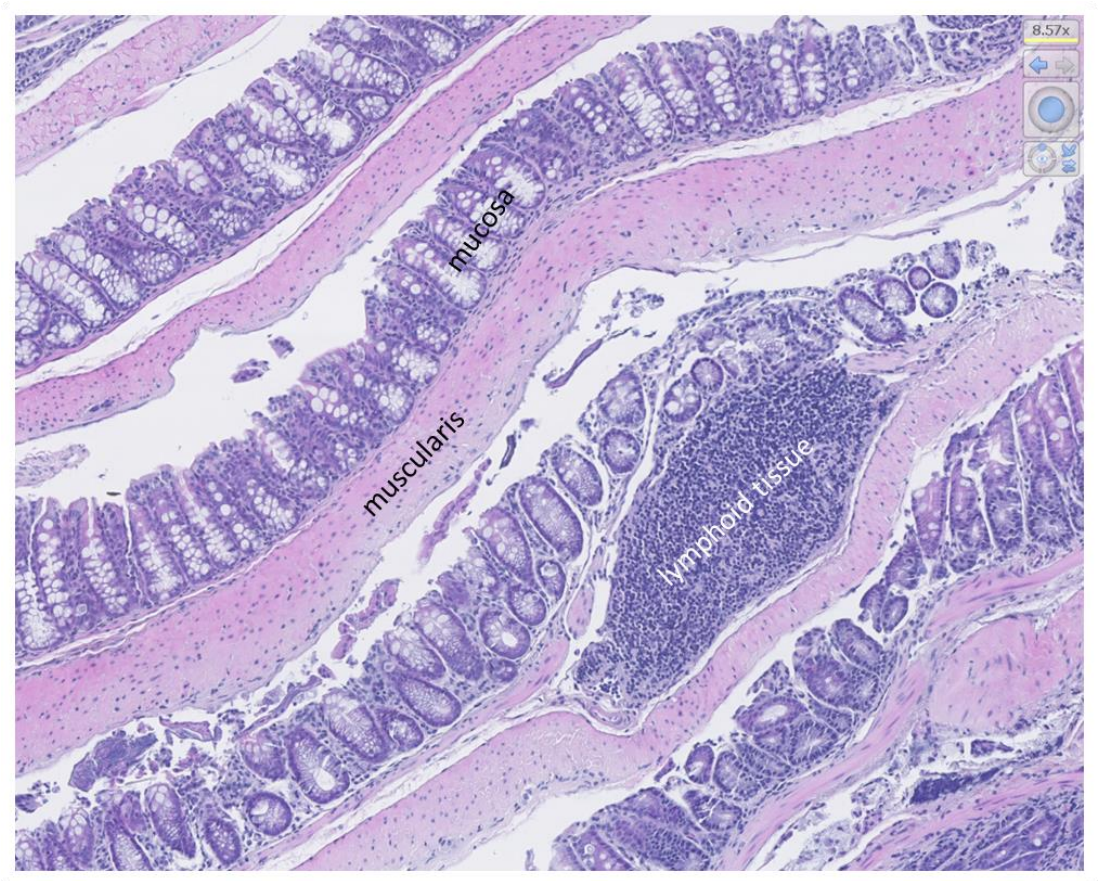


Figure 4.2: Section of fixed Swiss-rolled mouse large intestine, stained with haematoxylin and eosin (H&E). It is possible to distinguish the mucosal epithelium, the underlying muscularis propria layer, and occasional pockets of lymphoid tissue. Image taken at 85X.

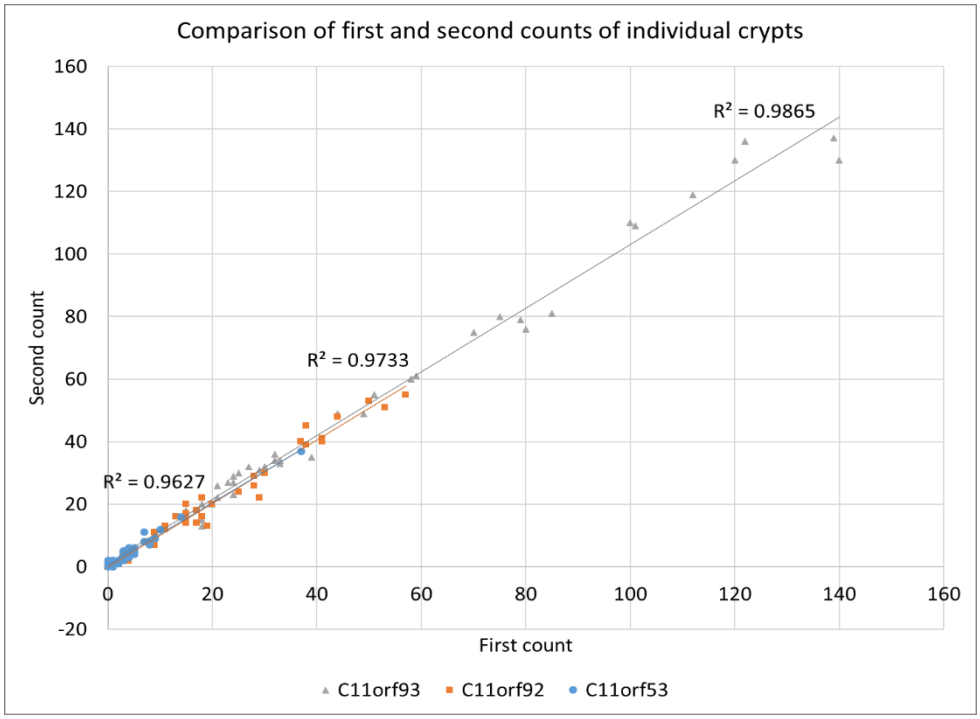


Figure 4.3: Comparison of initial count and recount for individual crypts within three human patients. I performed RNAscope on 10 human colon tissue samples for *C11orf53*, *C11orf92* and *C11orf93*, and counted the signal for each gene in each colonic crypt. Crypts were numbered so that each datapoint could be attached to a specific crypt on the slides. Three patients were randomly selected several months later and I recounted all of the RNAscope signal for each gene without reference to my original counts. I then compared the first and second counts. R^2 values are included for each gene, showing internal consistency in my identification and counting of RNAscope signals. The data shown is the total pooled results from across all three recounted patients.

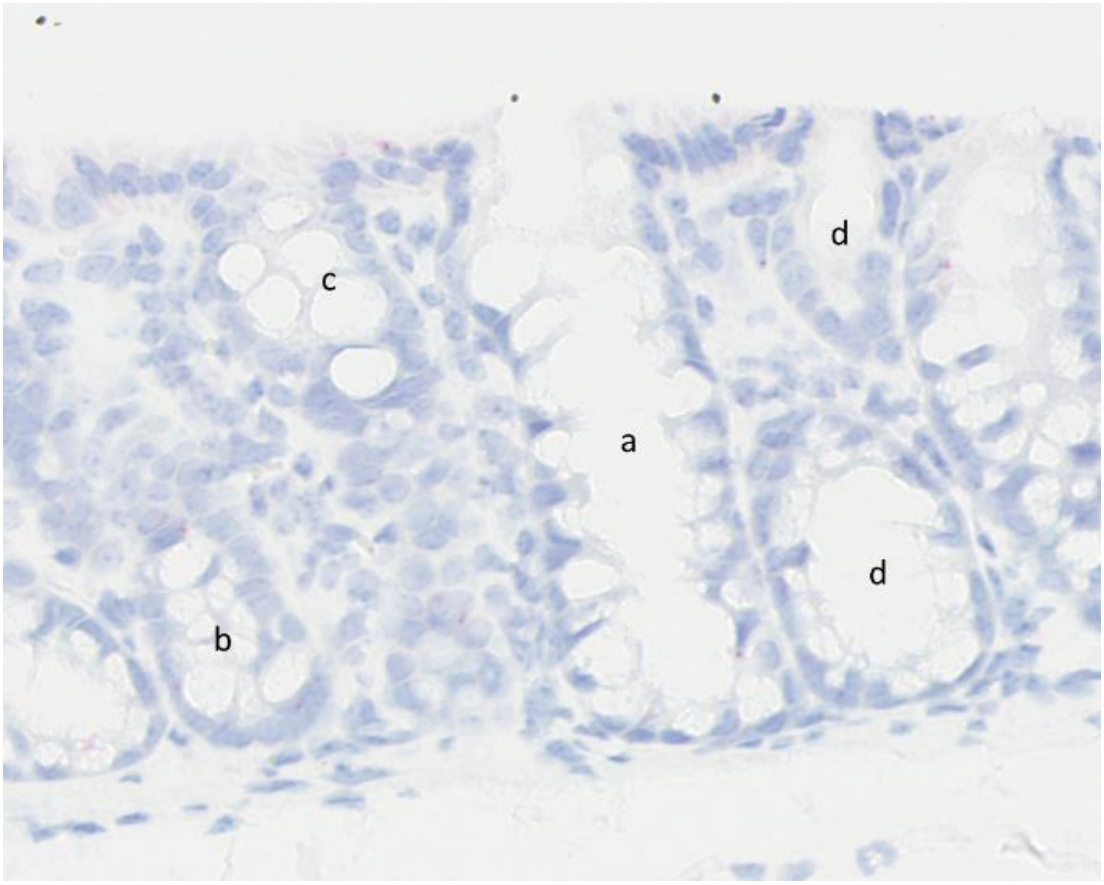


Figure 4.4: A section of WT mouse large intestine probed for *C11orf93*, showing (a) a properly sectioned crypt suitable for counting, and (b-d) improperly sectioned crypts that would not have been counted. Image taken at 350X.

4.2.2 Results

4.2.2.1 Expression patterns in normal intestinal mucosa

C11orf53 has the most distinct pattern of expression, occurring in compact clusters of transcripts with very little signal in between (figure 4.5a). *C11orf92* and *C11orf93* have more diffuse expression of individual dots (figure 4.5b, 4.5c). All three genes show highest expression towards the base of the crypt (figure 4.6). These patterns are consistent across human, WT mouse and $Apc^{min/+}$ mouse large intestine tissue. *C11orf93* has the highest expression in all three sample types, but in mice *C11orf53* and *C11orf92* have similar levels of expression, while in humans *C11orf92* is much higher than *C11orf53*. *C11orf53* is also more basal in humans than in mice, but *C11orf93* is less basal.

While I did not have any human small intestine tissue, I did analyse the crypts and the villi in WT and $Apc^{Min/+}$mouse small intestine, split into proximal and distal (figure 4.7a). As with the division of the crypts and villi into tertiles, this separation is based on a physical split rather than any biological marker; during collection and processing it is necessary to cut the small intestine into two lengths. In both proximal small intestine (PSI) and distal small intestine (DSI) *C11orf53* has particularly high expression in the villi, while *C11orf93* has the highest expression in the crypts. Comparing WT and $Apc^{Min/+}$, *C11orf53* is significantly lower in PSI crypts and villi and DSI crypts on the $Apc^{Min/+}$ background compared with WT. *C11orf92* is significantly lower in $Apc^{Min/+}$ PSI crypts but significantly higher in DSI crypts. *C11orf93* is significantly higher in all of the $Apc^{Min/+}$ SI crypts and villi, but not in the large intestine (LI) crypts.

I also compared the distribution of transcripts in the crypt and villi tertiles between WT and $Apc^{Min/+}$ (figure 4.7b). The region with the most difference is PSI crypts, where all three genes have significant changes in distribution: *C11orf53* has a reduction in expression in the top of the crypt and an increase in the middle, while *C11orf92* has a reduction in the base with increase in the middle and *C11orf93* has reduction in the base with increase in both the middle and top thirds. There are also significant differences with *C11orf53* in the PSI villi and *C11orf92* in the DSI crypts, in both cases with lower expression in the base tertile.

In addition to the crypts themselves, I counted the signal in the underlying muscularis and in patches of lymphoid tissue in the mice (figure 4.8). In the muscularis, which was measured in 500µm strips, *C11orf93* has the highest expression and *C11orf53* has the lowest expression.

However, *C11orf93* is significantly lower on the $Apc^{Min/+}$ background compared to WT, while *C11orf53* is significantly higher. *C11orf92* has much higher expression than the other two genes in lymphoid tissue, which was measured in 3000µm^2 regions. *C11orf53 and C11orf92* genes show significantly higher lymphoid expression in $Apc^{Min/+}$ mice compared to WT.

The human patients had previously been genotyped for rs3802842 and rs7130173. For both SNPs, patients who are homozygous for the risk alleles show significantly lower expression than those homozygous for the non-risk allele (figure 4.9a, 4.9b), as would be expected from the previous expression analysis (Smillie, 2015).

As well as the overall lower expression, risk alleles of the SNPs are associated with altered distribution of transcripts (figure 4.9c, 4.9d). All three genes have significantly higher proportions of their expression in the base third of the crypt in patients with the risk alleles. *C11orf53* and *C11orf93* have significantly lower proportions in the top third, while *C11orf92* has significantly lower proportion in the middle third.

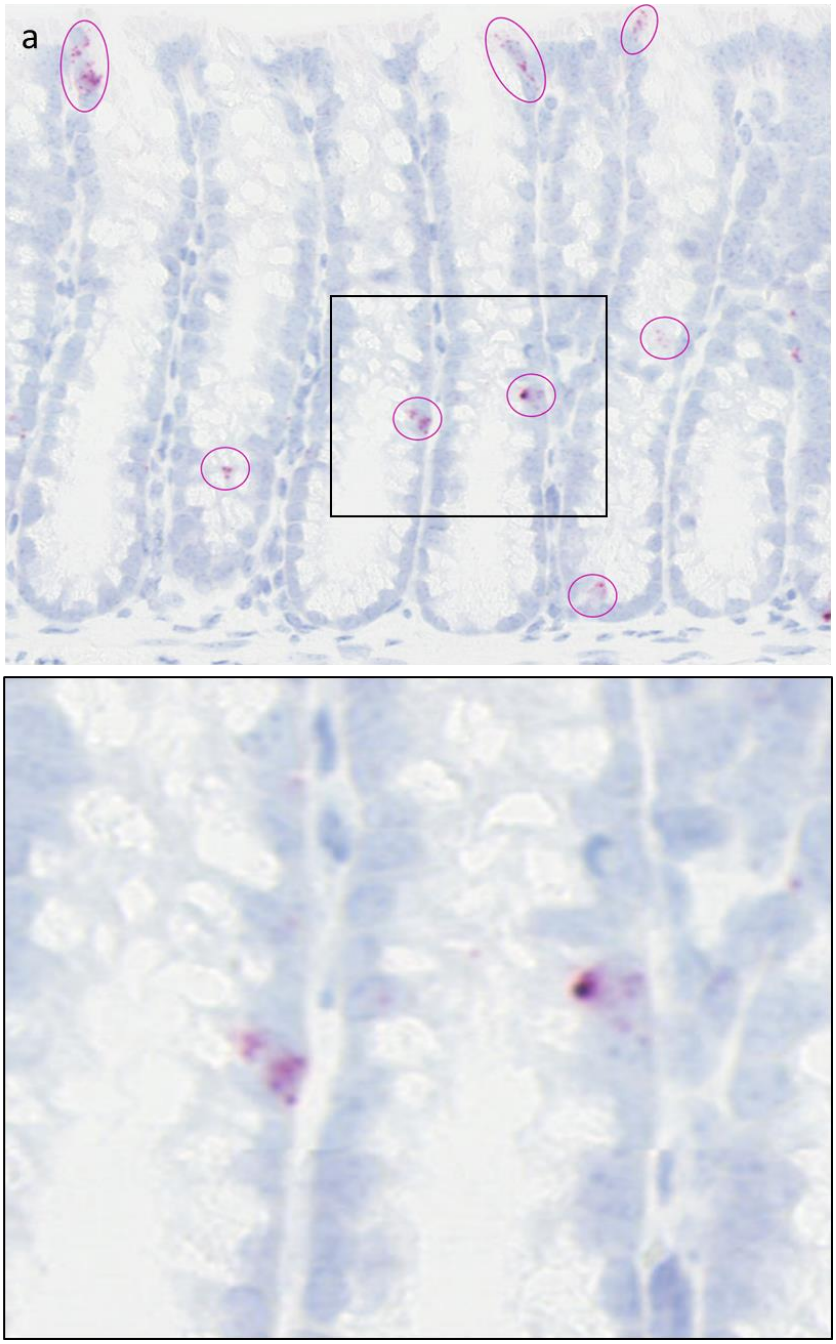


Figure 4.5a: RNAscope on WT mouse large intestine, probing for *C11orf53*. Dots have been circled to aid visualisation. Upper panel taken at 350X, with lower panel a further 3X magnification of the marked region.

For full RNAscope images (small and large intestine in WT and $Apc^{Min/+}$ mice and large intestine in humans) see appendix 1.

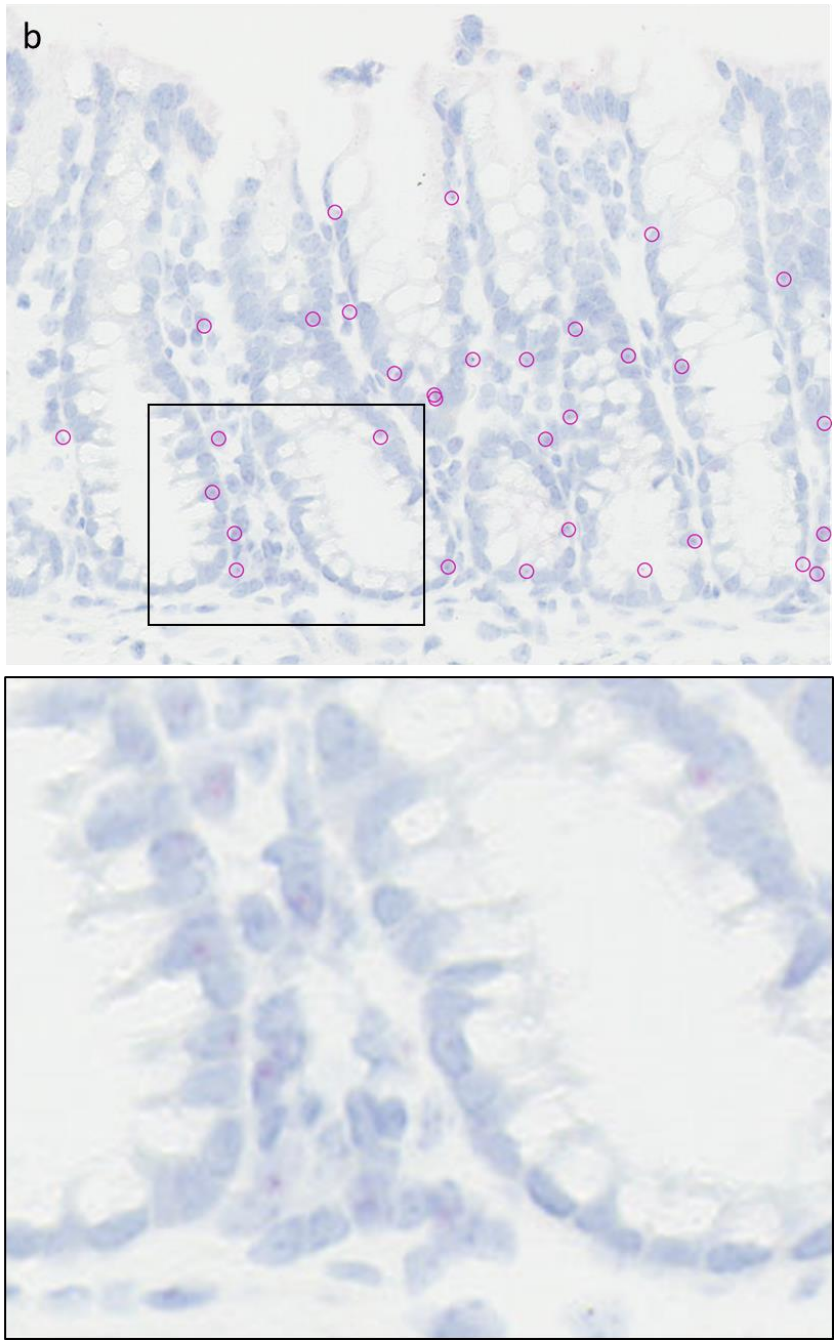


Figure 4.5b: RNAscope on WT mouse large intestine, probing for *C11orf92*. Dots have been circled to aid visualisation. Upper panel taken at 350X, with lower panel a further 3X magnification of the marked region.

For full RNAscope images (small and large intestine in WT and $Apc^{Min/+}$ mice and large intestine in humans) see appendix 1.

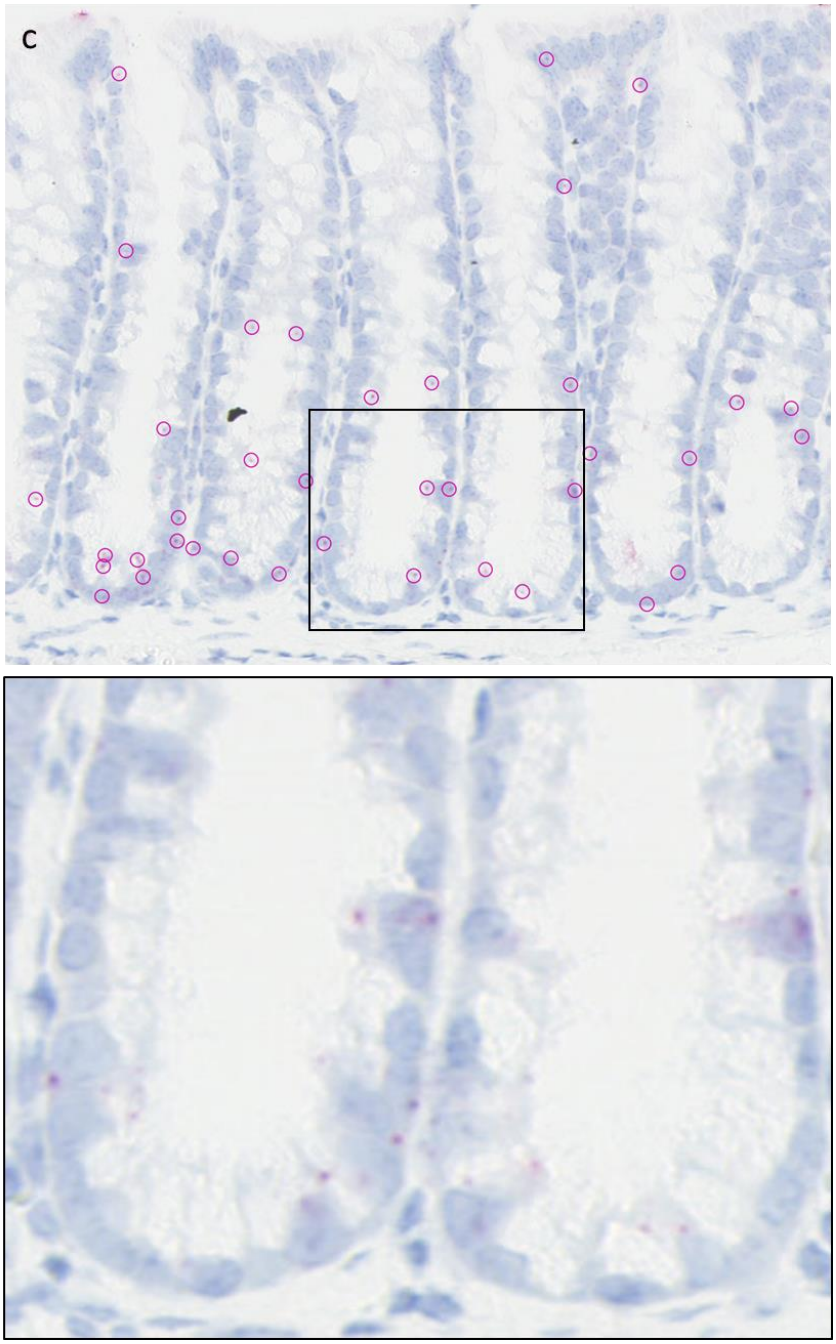


Figure 4.5c: RNAscope on WT mouse large intestine, probing for *C11orf93*. Dots have been circled to aid visualisation. Upper panel taken at 350X, with lower panel a further 3X magnification of the marked region.

For full RNAscope images (small and large intestine in WT and $Apc^{Min/+}$ mice and large intestine in humans) see appendix 1.

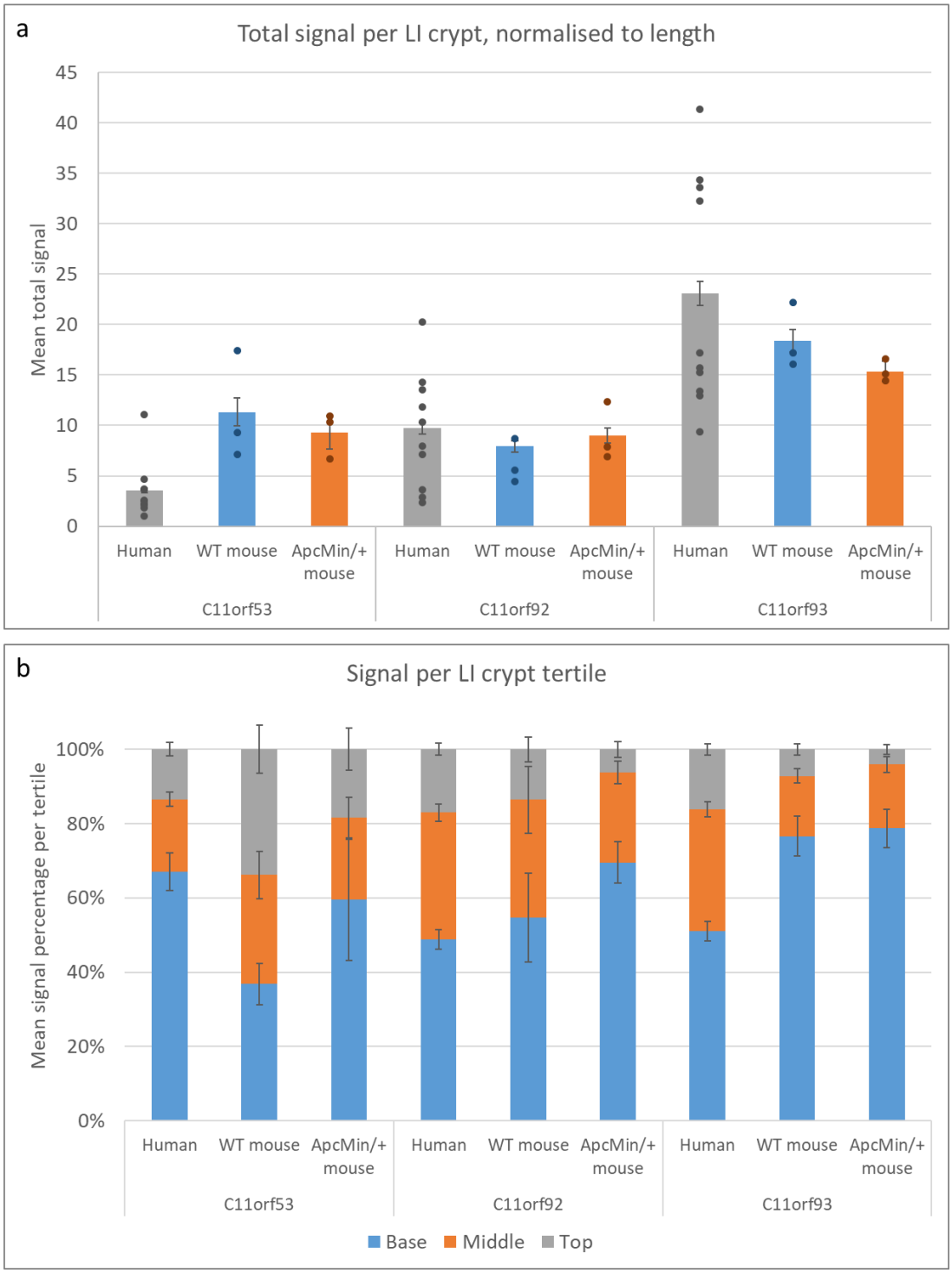


Figure 4.6: Mean expression of *C11orf53*, *C11orf92* and *C11orf93* as detected by RNAscope in LI crypts in human (n=10, mean 16.5 crypts per patient per gene), WT mouse (n=3, mean 19.1 crypts per animal per gene) and *Apc*$^{Min/+}$ mouse (n=3, mean 19.1 crypts per animal per gene), normalised to crypt length. (a) Mean total expression per LI crypt. Bars represent the mean of all crypts, with the mean for each human or mouse shown as an individual data point. (b) Mean percentage of expression in each LI crypt tertile.

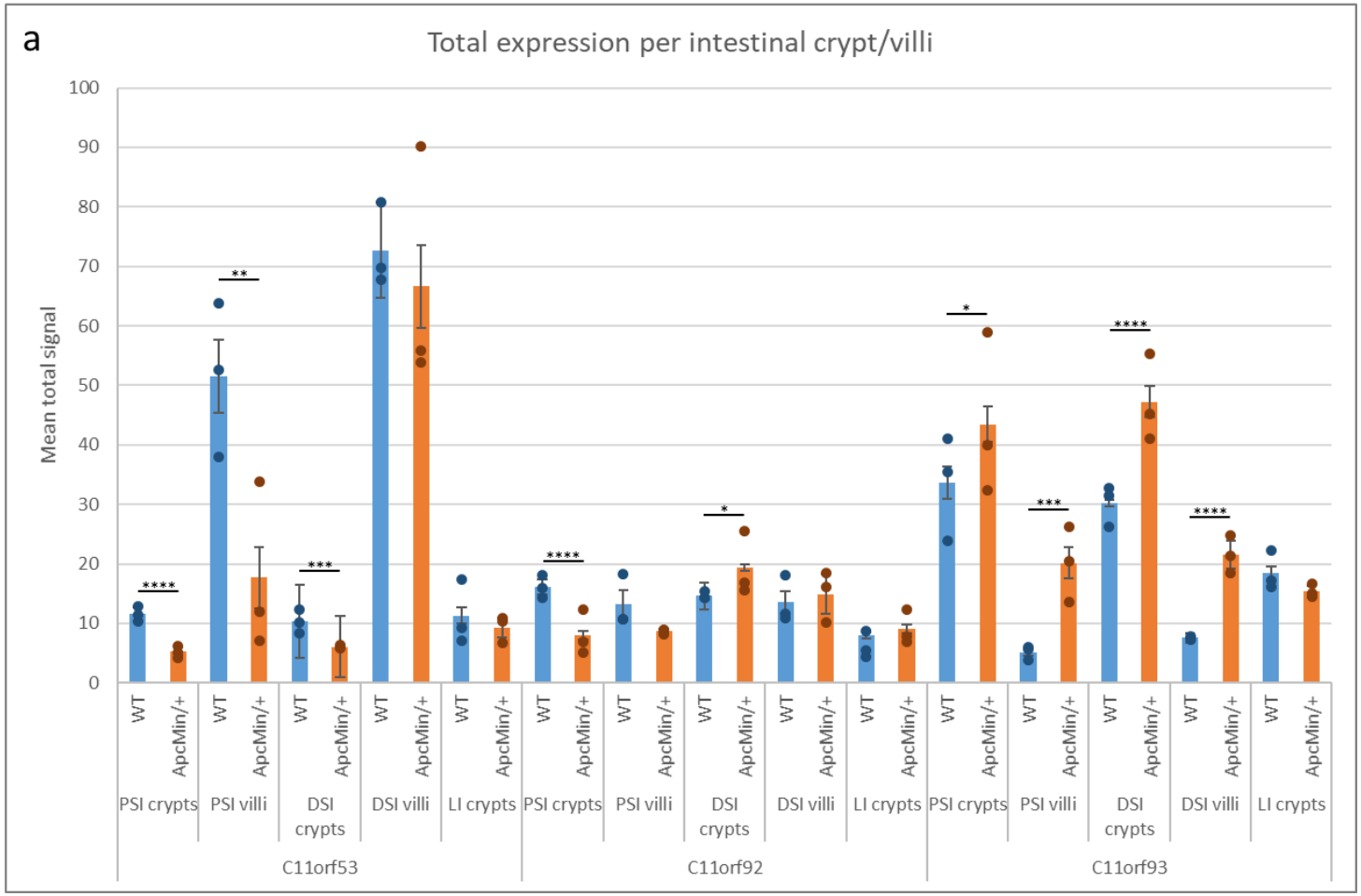


Figure 4.7a: Mean total expression of *C11orf53*, *C11orf92* and *C11orf93* in WT and $Apc^{Min/+}$ mice, as detected by RNAscope. n=3 animals per genotype, with mean 16.7 PSI crypts, 3 PSI villi, 16.7 DSI crypts, 3 DSI villi and 19.1 LI crypts per animal per gene. Expression normalised to length of crypt/villi. Bars represent the mean of all crypts/villi, with the mean for each animal shown as an individual data point. Data was a mix of normally and not-normally distributed according Shapiro-Wilk normality test, and so was compared using two-tailed Mann Whitney U test or student's t test as appropriate. Significant differences are indicated by stars.

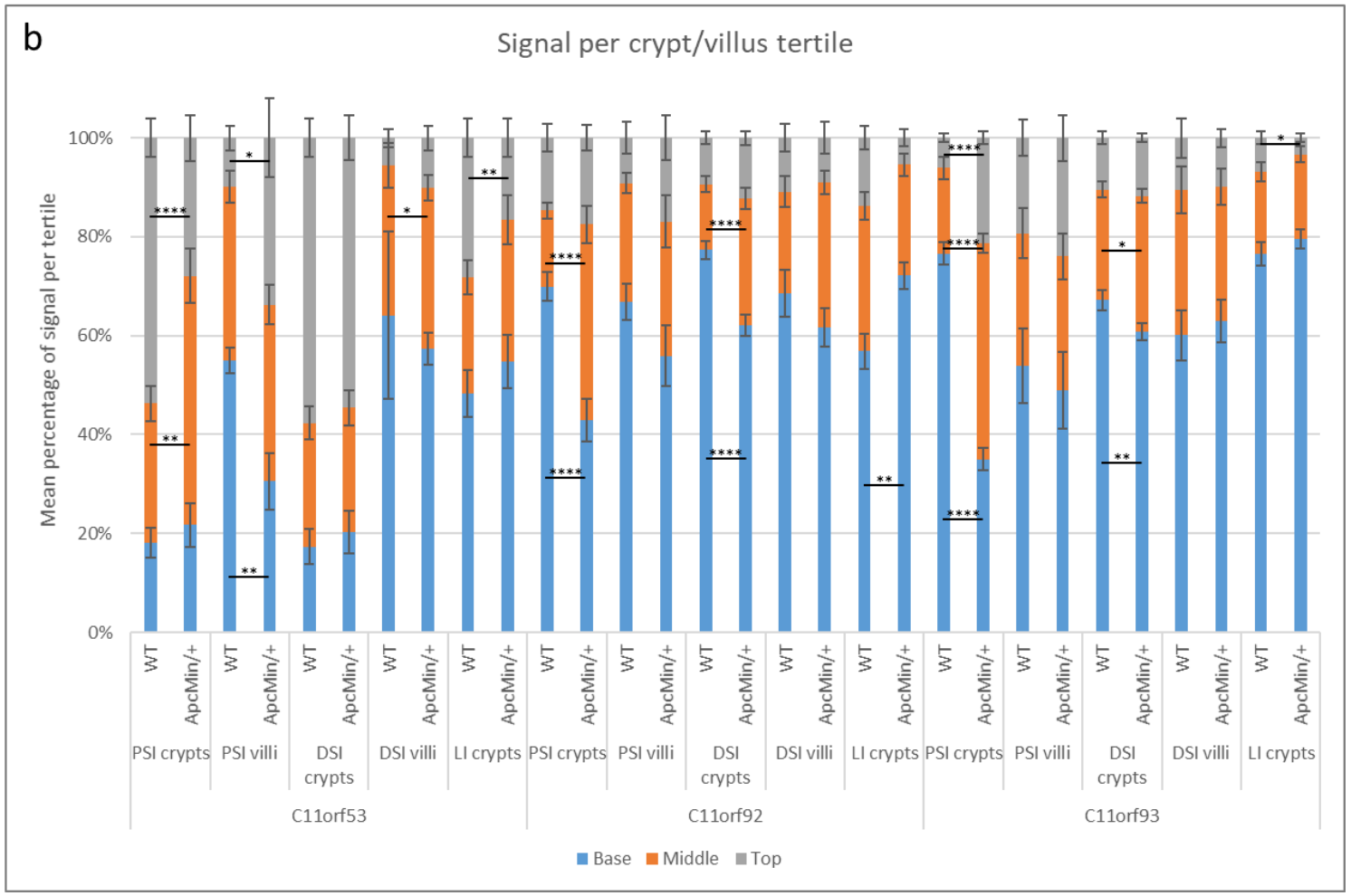


Figure 4.7b: Mean percentage of expression of *C11orf53*, *C11orf92* and *C11orf93* in each crypt or villus tertile in WT and $Apc^{Min/+}$ mice, as detected by RNAscope. n=3 animals per genotype, with mean 16.7 PSI crypts, 3 PSI villi, 16.7 DSI crypts, 3 DSI villi and 19.1 LI crypts per animal per gene. Data was a mix of normally and not-normally distributed according Shapiro-Wilk normality test, and so was compared using two-tailed Mann Whitney U test or student's t test as appropriate. Significant differences are indicated by stars.

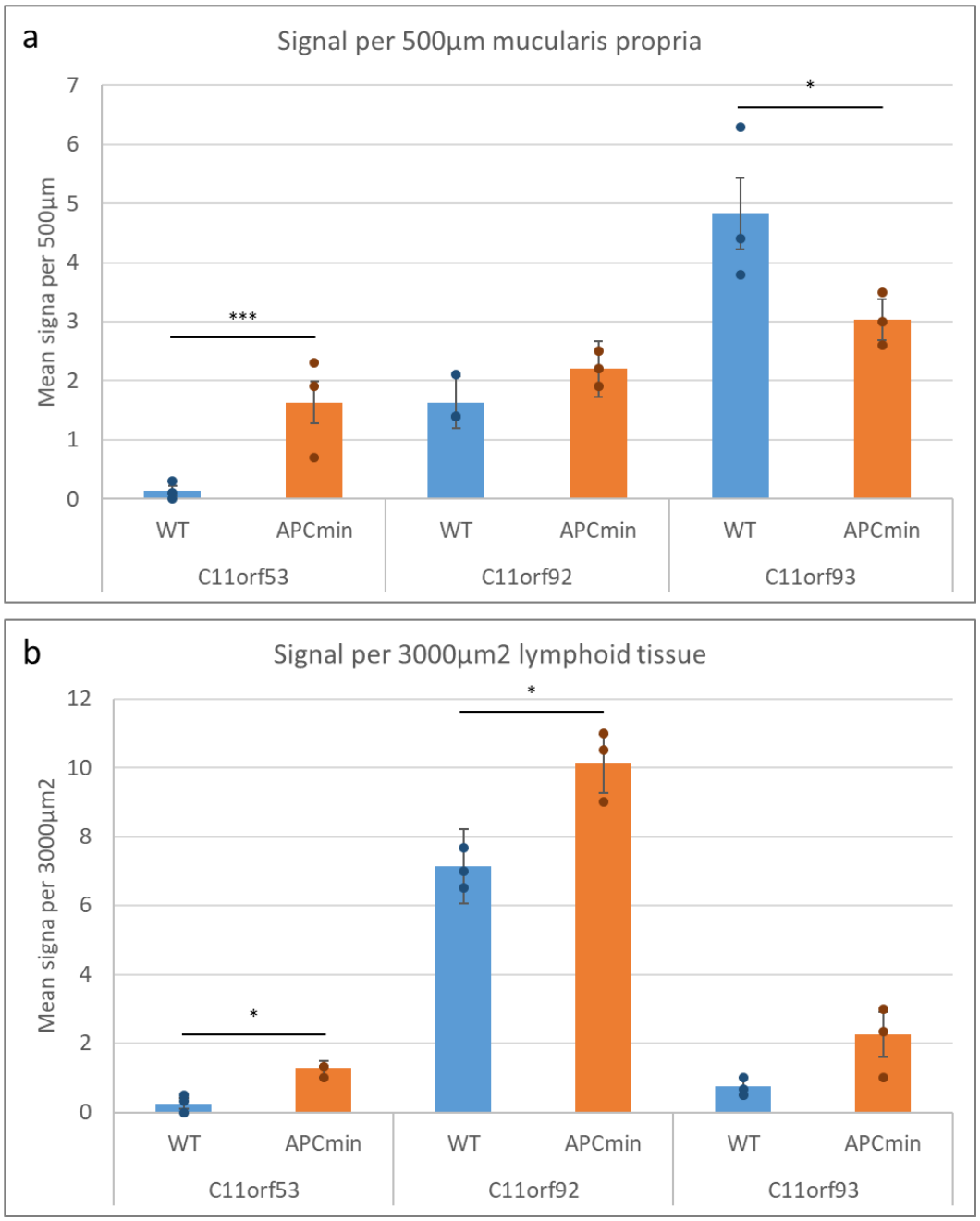


Figure 4.8: Mean expression of *C11orf52*, *C11orf92* and *C11orf93* in WT and $Apc^{Min/+}$ mice as detected by RNAscope per (a) 500µm large intestine muscularis propria, n=3 animals per genotype, mean 10 sections of muscularis propria per animal; (b) 3000µm^2 large intestine lymphoid tissue, n=3 animals per genotype, mean 2.7 sections of lymphoid tissue per animal. Bars represent the mean of all sections, with the mean for each animal shown as an individual data point. Data was a mix of normally and not-normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test or student's t test as appropriate. Significant differences are indicated by stars.

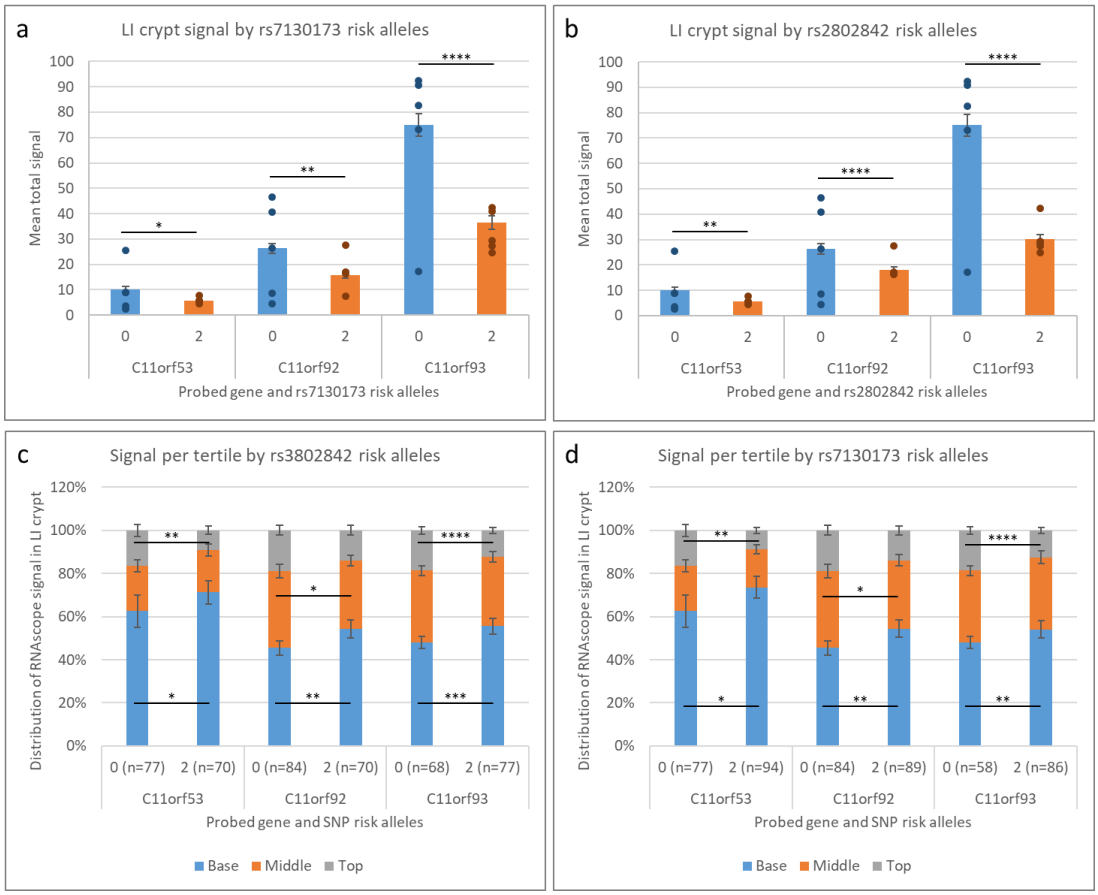


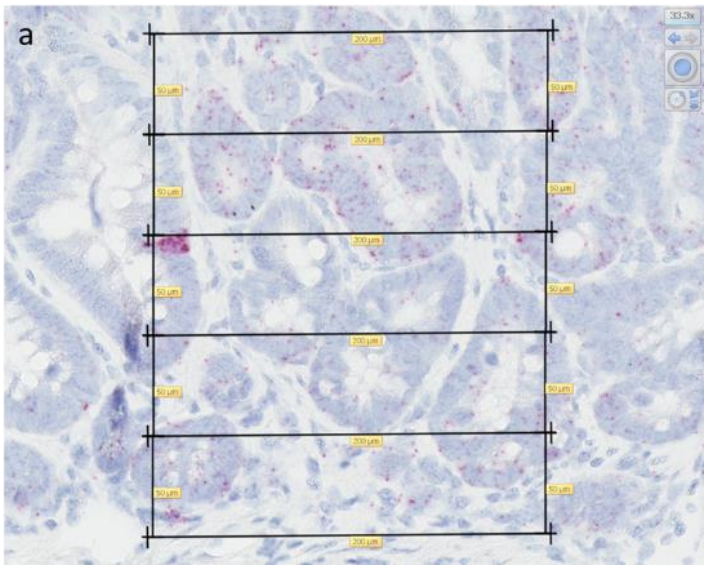
Figure 4.9: Differences in expression between SNP genotypes as detected by RNAscope. (a) Mean total signal per LI crypt by risk alleles of tagging SNP rs3802842. For 0 risk alleles, n=5, with mean 15.3 crypts per patient per gene. For 2 risk alleles, n=4, with mean 18.1 crypts per patient per gene. (b) Mean total signal per LI crypt by risk alleles of probable causal SNP rs7130173. For 0 risk alleles, n=5, with mean 15.3 crypts per patient per gene. For 2 risk alleles, n=5, with mean 17.9 crypts per patient per gene. (c) Mean RNAscope signal distribution between LI crypt tertiles by risk alleles of rs3802842, using same data as (a). (d) Mean RNAscope signal distribution between LI crypt tertiles by risk alleles of rs7130173, using same data as (b). Bars represent the mean of all sections, with the mean for each patient shown as an individual data point. Data was a mix of normally and not-normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test or student's t test as appropriate. Significant differences are indicated by stars.

4.2.2.2 Expression patterns in intestinal polyps and tumours

The crypts analysed in the *Apc*$^{Min/+}$ mice and human patients were all from normal mucosa, despite the presence of growths in their intestines. In the humans, the normal mucosa was sourced from a separate site than that of the tumour, and in the mice I made sure to leave a "buffer" of ~5 crypts either side of the visible edges of polyps.

However, I examined the mouse polyps separately. The architecture is greatly disrupted so crypts are no longer useful structures for counting. Instead, I split the polyps into strips of set dimensions, 200x50µm, starting at the base of the epithelium and progressing into the polyp (figure 4.10). There is no pattern of expression either increasing or decreasing with polyp depth. There is also much variation between different polyps. Basic expression patterns seen in normal mucosa are maintained in polyps: *C11orf93* has the highest expression and *C11orf53* the lowest, and *C11orf53* occurs in clusters while *C11orf92* and *C11orf93* are diffuse.

This conservation of transcript distribution is also seen in human tumours. Patients that have low expression in normal mucosa have minimal expression in tumours, even lower than in normal mucosa, while those with high expression in normal mucosa have similar or higher expression in tumours. The pattern of clustered or diffuse is also maintained.



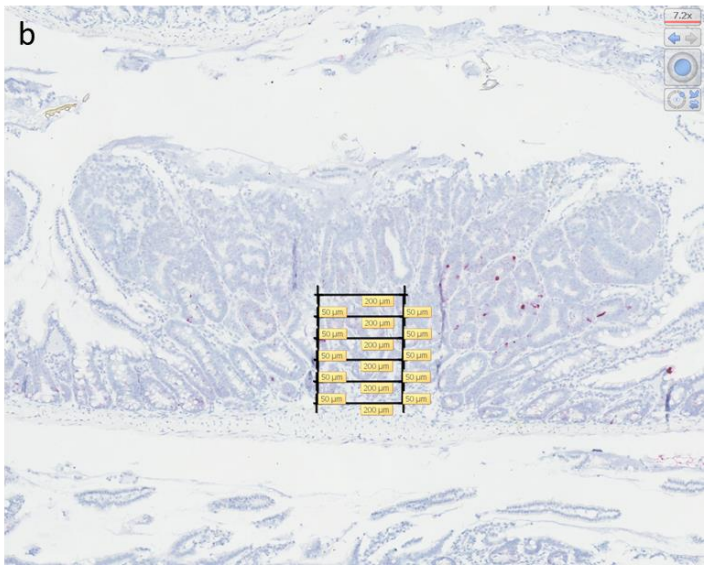
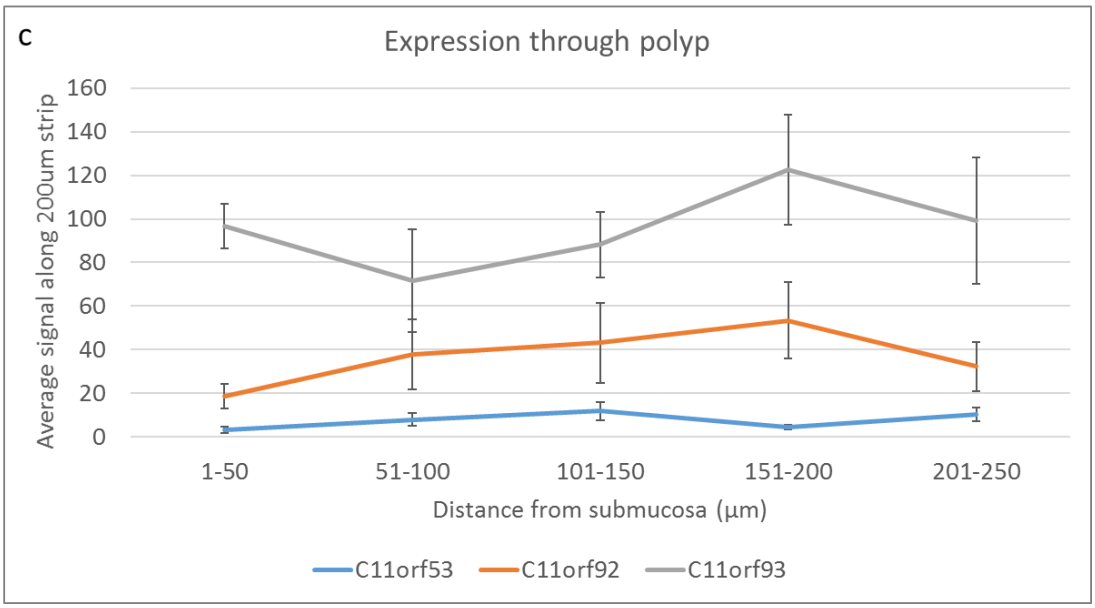


Figure 4.10: (a) *Apc*$^{Min/+}$ polyp that has been probed for *C11orf93* and split into sections 200x50µm, beginning with the base of the crypts at the submucosa and extending into the polyp. Image taken at 333X. (b) Position of the sections within the whole polyp. Image taken at 72X. (c) Mean expression of *C11orf* genes in each strip of polyp, as shown in (a) and (b), with increasing distance from the base of the crypts. There is no pattern of increase or decrease in gene expression with increased distance into the polyp. Data shown is from six polyps within one mouse.



4.2.2.3 Comparison with qRT-PCR and HT12 microarray results

When the human patients were undergoing surgery and tissue collection, in addition to the normal mucosa sections that were processed into FFPE blocks, mRNA was also extracted from fresh tissue. Expression analysis was carried out by Claire Smillie and Li-Yin Ooi (Ooi, 2016; Smillie, 2015). Patients were classified as either "high expressors" or "low expressors" based on qRT-PCR results. Illumina HT12 microarrays were also run on the patients. My ten patient samples comprised five high expressors and five low expressors, selected for me by another member of the group who could see their expressor status, while I was blinded to this information. Following RNAscope data collection on these patients, I then tried to identify the two groups of five. I was able to correctly assign eight of the ten patients (figure 4.11).

All ten patients had qRT-PCR results and eight also had HT12 microarray results. Once I was unblinded and given access to this data, I plotted each patient's mean RNAscope score against their qRT-PCR score, and HT12 microarray score if available (figure 4.12).

The correlation between RNAscope and these other RNA-based techniques is not very high. One reason for this weak correlation may be the nature of the probes for each technique (figure 4.13). RNAscope uses a pool of ~20 probe pairs across a wide region of the transcript, and only requires three to bind to give a signal. This is to ensure a signal can still be obtained even if sections of the molecule are degraded or concealed by bound proteins. qRT-PCR and HT12 microarrays use a single probe, or pair of probes, so if this section of sequence is damaged no product can be generated. The extent to which this may occur will vary between samples and genes, giving the lower correlation.

Another reason might be due to cell types present in the source of the expression data. As RNAscope is performed on fixed tissue and the signal assessed within the tissue context, it is possible to be very exacting in only counting transcripts within the mucosal layer, and excluding anything in the muscularis or other tissue type. The RNA used for qRT-PCR and HT12 analyses was extracted from stripped mucosa, and was shown to be enriched for epithelial markers, but the samples will still not have been a completely pure mucosal population.

C11orf53 has particularly poor correlation between RNAscope and qRT-PCR. It is possible that, because the signal clusters, I undercount higher expression due to not being able to

fully distinguish transcripts when they are at high density. *C11orf53* has the highest correlation between HT12 microarray and RNAscope, but this is primarily driven by a single patient with a particularly high value for both.

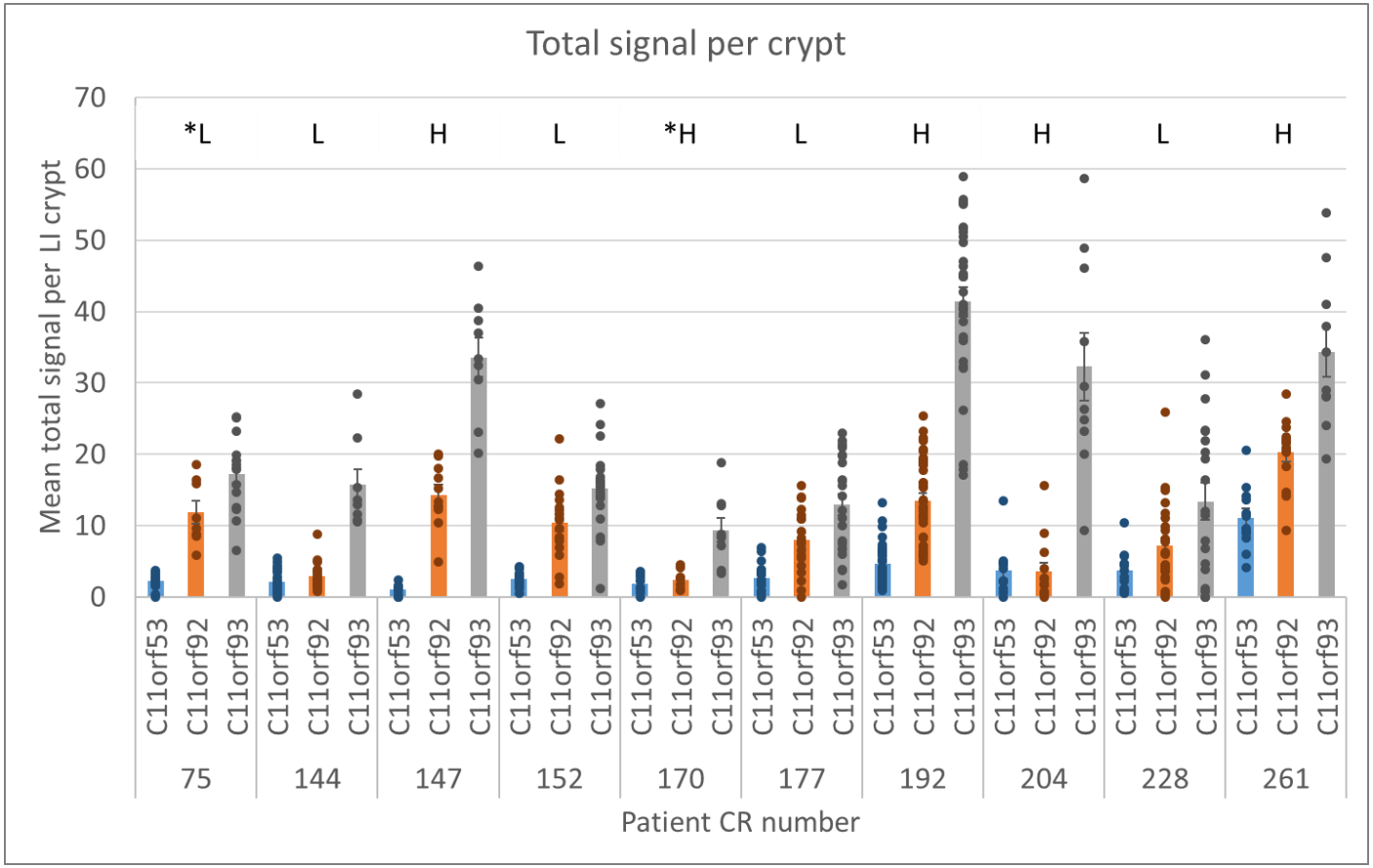


Figure 4.11: Mean RNAscope expression of each *C11orf* gene per LI crypt in 10 patients. Bars represent the mean of all crypts, with each crypt shown as an individual data point. Letter underneath the CR number indicates whether they were designated a high (H) or low (L) expressor, based on their qRT-PCR results. Asterisks mark patients I was not able to correctly identify as such from their RNAscope results. Misidentification of patient CR75 was primarily due to their *C11orf92* expression, as their results for *C11orf53* and *C11orf93* are consistent with other low expressors. Patient CR204 also has *C11orf92* expression that appears out of proportion with their *C11orf93* result and their classification of being a high expressor; as there is little variation between *C11orf53* expression between high and low expressors, this patient's classification of high for the locus as a whole may be being driven purely by their high *C11orf93* expression

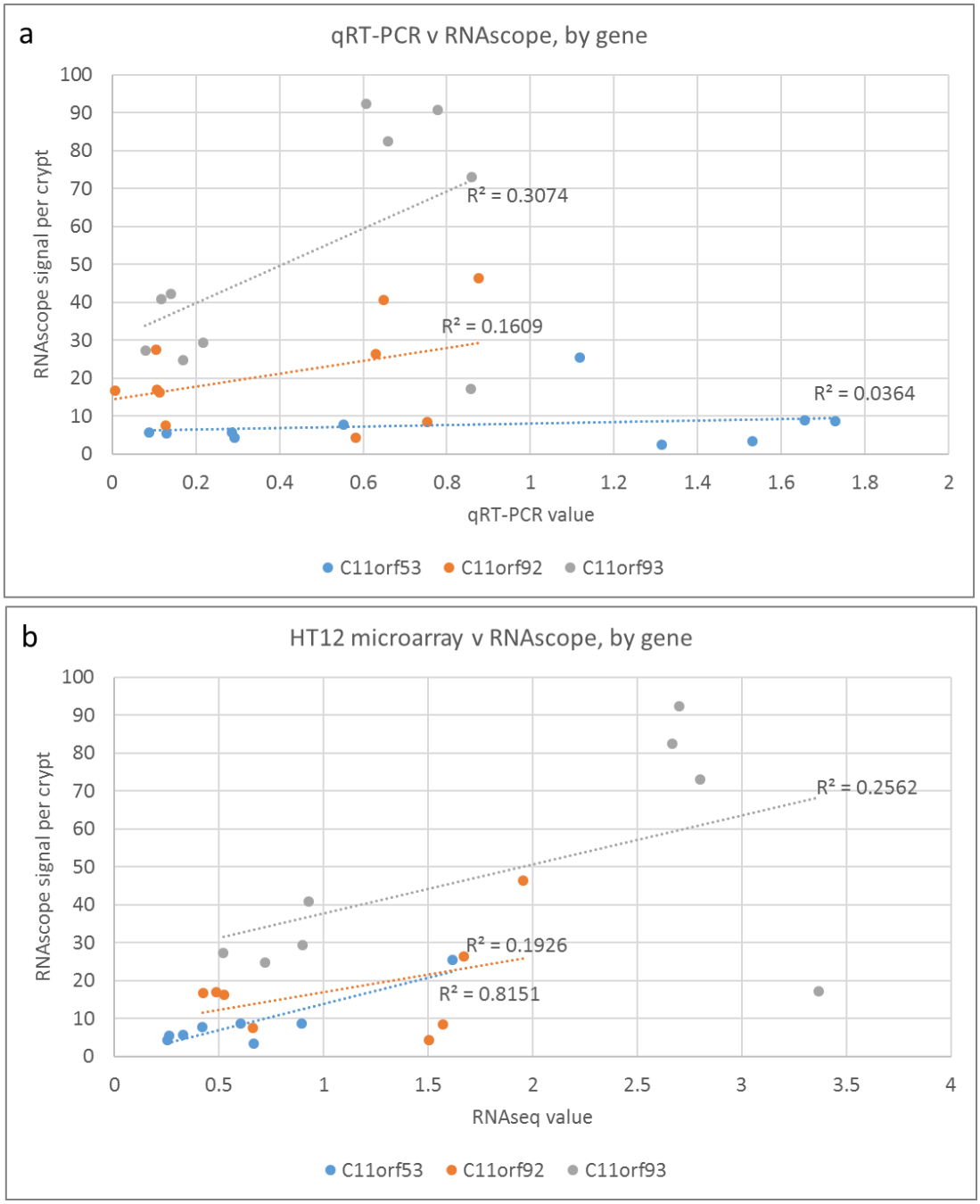


Figure 4.12: (a) Mean RNAscope signal for *C11orf53*, *C11orf92* and *C11orf93* per large intestine crypt for 10 patients plotted against their qRT-PCR value for each gene, which was generated from RNA extracted from stripped mucosal tissue. Trendlines and R^2 values are included. (b) Mean RNAscope signal for *C11orf53*, *C11orf92* and *C11orf93* per large intestine crypt for 8 patients plotted against their Illumina HT12 microarray value for each gene, which was generated from RNA extracted from stripped mucosal tissue. Trendlines and R^2 values are included.

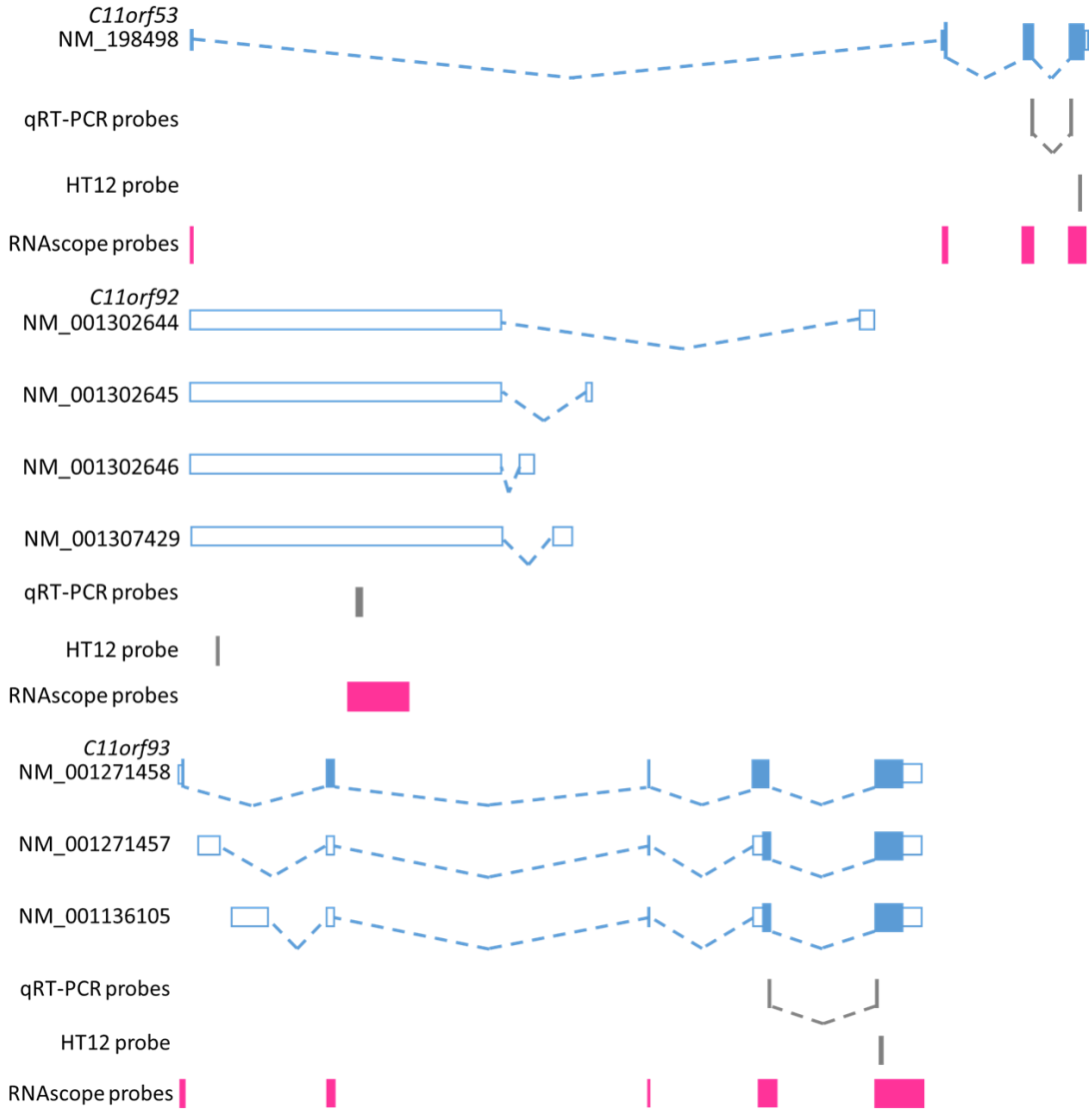


Figure 4.13: Schematics of the known transcripts of *C11orf53*, *C11orf92* and *C11orf93*, in blue, showing the locations of qRT-PCR and HT12 probes in grey, and RNAscope probes in pink. While the qRT-PCR and HT12 probes detect only small regions of the mRNA, the RNAscope system of a pool of up to 20 probe pairs allows coverage of a much larger proportion of the transcript, so giving greater sensitivity in circumstances where the mRNA may be partially degraded or partially coated in proteins.

.4.2.2.4 Co-expression between *C11orf* genes and other markers

As *C11orf92* and *C11orf93* have similar expression patterns in their distribution within the crypt, we checked to see if they are being co-expressed. Serial sections were cut from WT mouse tissue blocks and I probed alternate sections for the two genes. I then found crypts that were well orientated on both slides and overlaid the signal from one probe onto the other (figure 4.14). These overlays do not suggest the two genes are only expressed together.

The stem cells in the intestine are located in the base of the crypts and Lgr5 is used as a key marker of these stem cells (Barker et al., 2007). As the *C11orf* genes show the highest expression in base of the crypt, I carried out 2-plex RNAscope on WT mouse tissue to see how much co-expression there is between these genes and *Lgr5* (figure 4.15). While the genes are not exclusively expressed – there are cells that are both $C11orf^{+}$ and $Lgr5^{+}$ - there are also cells that only express one of the two.

It is possible to couple RNAscope to an antibody-based technique, either immunohistochemistry or immunofluorescence. Although 2-plex RNAscope was the best option for *Lgr5*, as the antibodies available for Lgr5 are not very specific, this would be very useful for other markers in further investigation into which cell types do or do not express the *C11orf* genes. Unfortunately so far we have been unsuccessful at carrying this out in our lab – combining the pre-treatment conditions for the technique has so far resulted in either under-treated tissue in which the antibodies cannot access the antigens, or over-treated tissue in which structural definition is lost (figure 4.16).

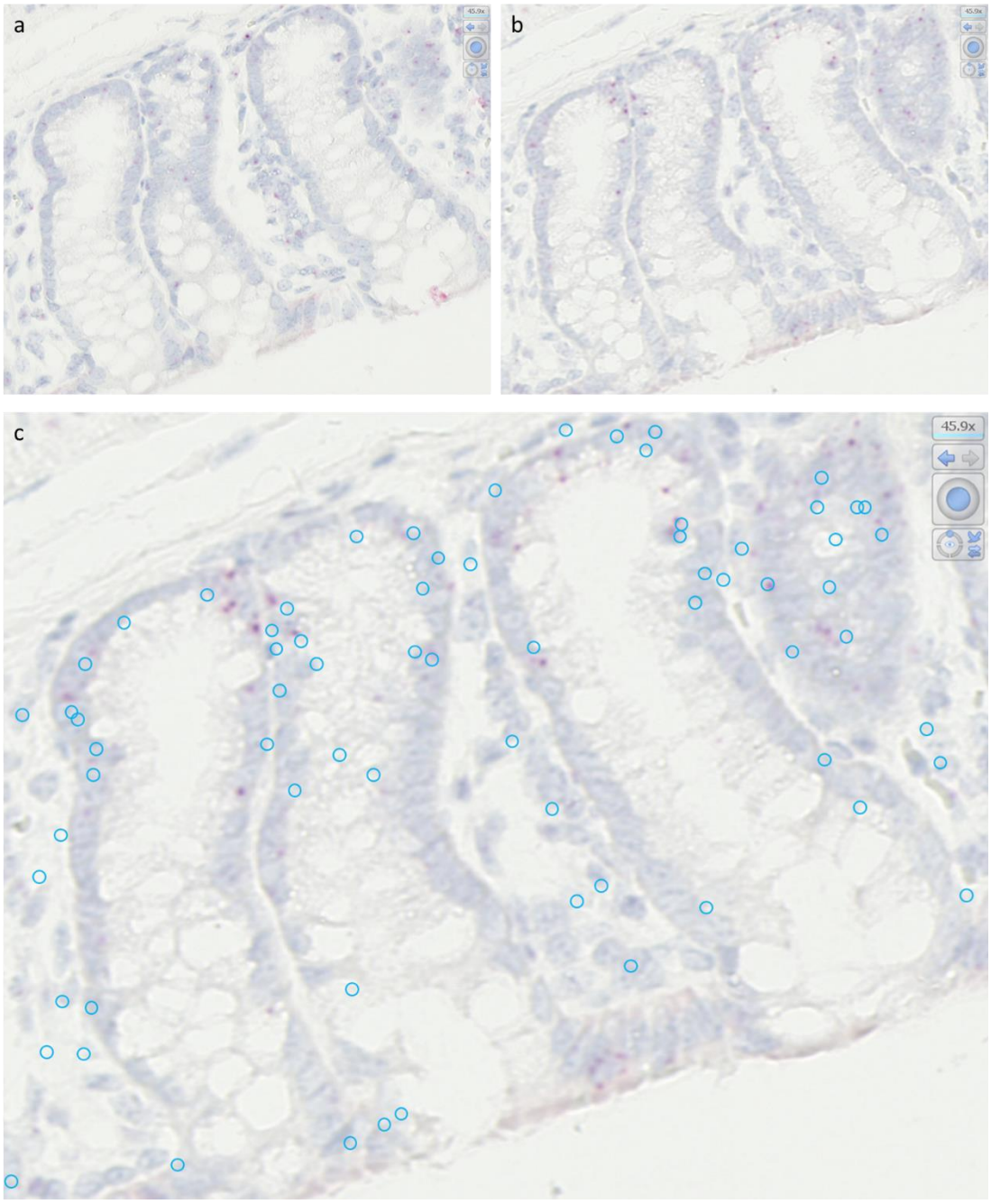


Figure 4.14: Serial sections of WT mouse LI, probed for (a) *C11orf92* and (b) *C11orf93*. (c) The location of the *C11orf92* signal, represented by blue circles, is overlaid onto the image of *C11orf93* signal, seen as pink dots. The two genes show clearly different expression patterns in the two adjacent 5µm tissue sections. Images taken at 460X.

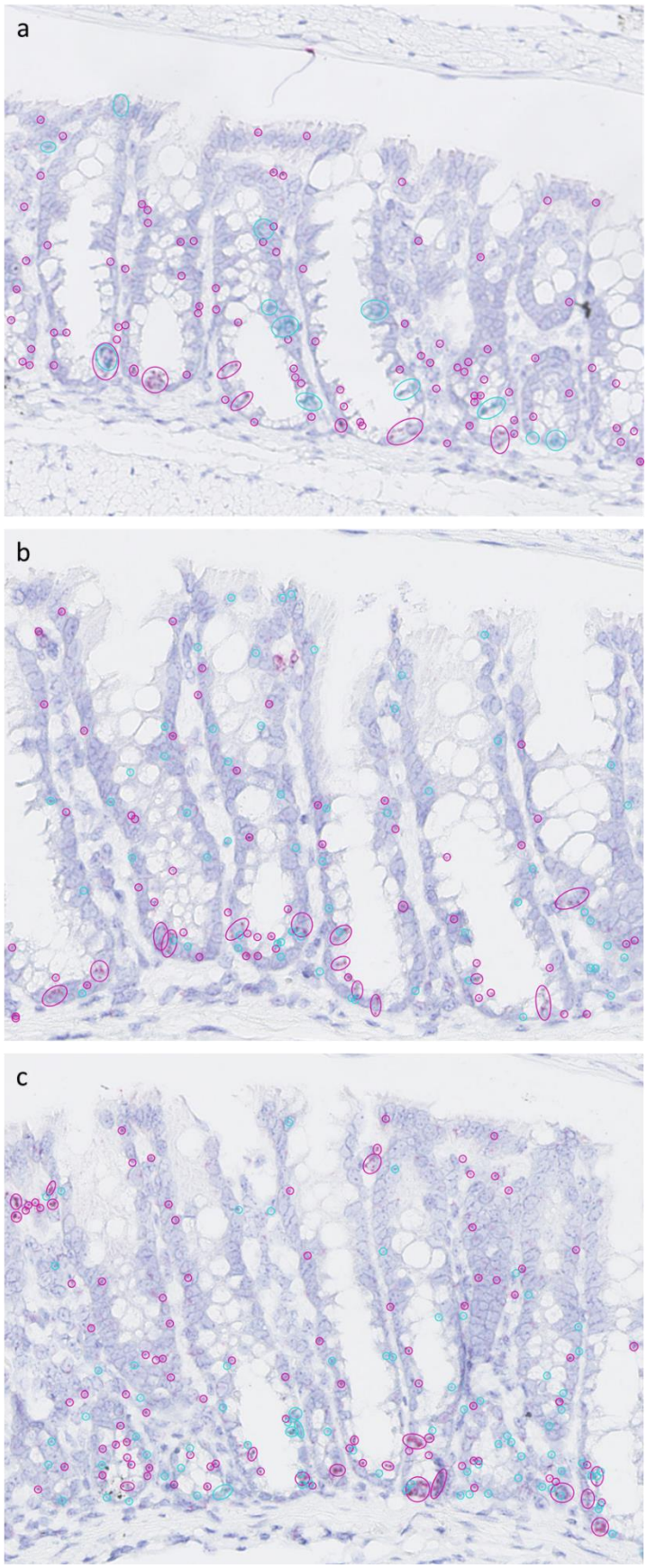


Figure 4.15: Wild type mouse large intestine that has been treated with duplex RNAscope to visualise expression patterns between multiple genes on the same tissue section. *Lgr5*, a marker for intestinal stem cells, is detected with pink probe. (a) *C11orf53*, (b) *C11orf92* and (c) *C11orf93* are detected with teal probe. Signal has been circled to aid visualisation. Images taken at 300X.

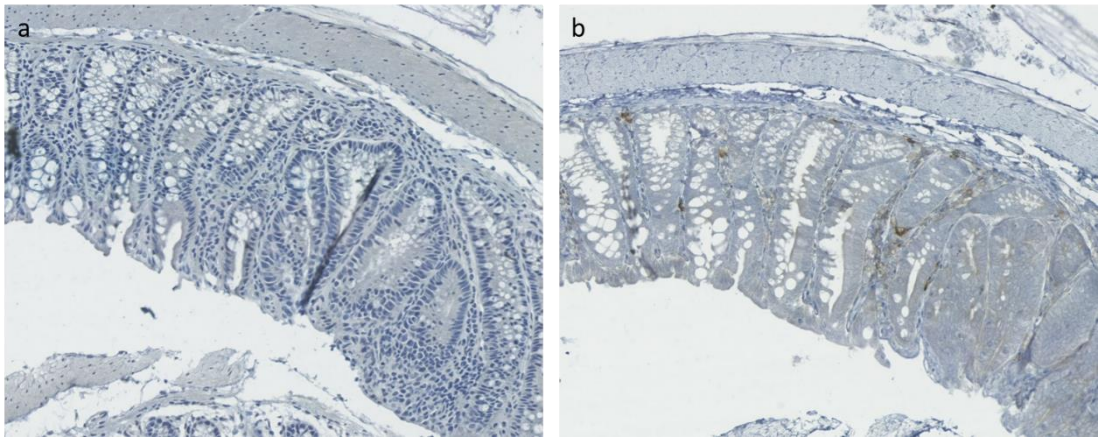


Figure 4.16: Examples of our attempts to combine RNAscope with immunohistochemistry (IHC), staining for β-catenin. (a) The tissue was pre-treated using the only the conditions for RNAscope, proceeding directly from RNAscope amplification to antibody incubations. The tissue has been undertreated for IHC and the antibodies cannot access antigens, giving no staining. (b) The tissue was pre-treated for RNAscope, and further pre-treatment for IHC was performed between RNAscope amplification and antibody incubation. The tissue has been overtreated, destroying structural definition. RNAscope must be performed before IHC as RNA is less likely to survive IHC than proteins are to survive RNAscope, in part because antibodies are often supplied in solution with nucleases.

4.3 Discussion

The use of in situ RNA hybridisation has allowed us to examine the expression patterns of the *C11orf* genes, despite the lack of useable antibodies. This provides a range of new information on their activity, some of which lends support to previous conclusions and some of which allows novel ideas to be developed.

It is clear, from looking at all of the expression data gathered from across these tissues, that *C11orf53*, *C11orf92* and *C11orf93* each have their own distinct expression pattern. This further supports the view that this locus is in fact three separate genes, rather than a single open reading frame called *FLJ45803*, as was once thought.

It is also clear that these genes show consistent patterns across human and mouse. The genes have different degrees of conservation between the two species – *C11orf53* is 83% conserved, *C11orf92* is 50% conserved and *C11orf93* is 65% conserved. As we have made extensive use of mouse models in the study of these genes, it is reassuring to have this confirmation that their behaviour, beyond the gene sequences, is conserved between species. It is also interesting to note that the gene with the lowest sequence conservation, *C11orf92*, has the highest similarity in transcript level and distribution between human and mouse.

C11orf53's pattern of expression strongly suggests it is being expressed in a specific cell type. There is little overlap between *C11orf53*$^+$ cells and *Lgr5*$^+$ cells, but some cells expressing both *C11orf53* and *Lgr5* have been observed, indicating that *C11orf53* is primarily expressed in differentiated cells rather than stem cells, but it is not exclusive to differentiated cells. Likewise, they are in the wrong place to be Paneth cells, also found in the base with the stem cells (Sato et al., 2011). They lack the distinctive "goblet" shape which gives goblet cells their name, caused by apical accumulation of secretory granules (Specian and Oliver, 1991). Other possible cell types include tuft cells and enteroendocrine cells. Being secretory, both these cell types must have extensive endoplasmic reticulum activity to process all the peptides to be secreted.

The overexpression immunofluorescence data showed that C11ORF53-DDK co-localises with calnexin, an endoplasmic reticulum protein (see chapter 3 for immunofluorescence data); furthermore, the pattern of *C11orf53* transcripts within the cells does itself somewhat resemble the endoplasmic reticulum. It is therefore reasonable to suggest that *C11orf53*$^+$

cells are some form of secretory cell, although determining which one, or ones, will require more investigation into co-expression with type-specific markers such as Dclk1 for tuft cells (Gerbe et al., 2012) and Cldn4 for enteroendocrine cells (Nagatake et al., 2014).

By contrast, *C11orf92* and *C11orf93* do not appear to be cell type-specific, being expressed across a higher number of cells along the length of the crypts. While the *Lgr5* 2-plex shows that these genes are not solely expressed in the stem cells, they do show higher expression in the region of the crypt where the stem cells are found, and this proportion increases both in $Apc^{Min/+}$ mice compared to WT mice, and in humans with more risk alleles of the identified risk SNPs rs3802842 and rs7130173, despite the accompanying overall reduction in expression level. This may suggest that, although they do have functions elsewhere in the colonic tissue, it is their activity in the stem cell region that gives their link to cancer initiation.

Since these genes are thought to be tumour suppressors, it might be anticipated that they would show reduced expression in the $Apc^{Min/+}$ model of tumourigenesis compared to WT, but there is no consistent pattern of this. *C11orf53* is lower throughout the intestines, but *C11orf93* is higher in the SI and *C11orf92* is higher in both DSI and LI.

This is unexpected, but not concerning. All of the eQTL data suggesting these genes are tumour suppressors was based on the LI, and they have not been implicated in the risk of any other types of cancer, including small intestinal, so we should not expect them to act as tumour suppressors in the SI. Additionally, the original GWAS and subsequent expression analysis were performed on patients with sporadic cancers, rather than FAP, which is the type of CRC the $Apc^{Min/+}$ mouse models. It has been shown that the tagging SNP rs3802842 does not affect colonic polyposis in FAP patients as it does in non-FAP patients (Cheng et al., 2015). This could be because the effect of the *Apc* mutation is so powerful it overwhelms the impact of lower-effect risk genes, but it could also be because these genes do not have the same tumour suppressor activity in the context of FAP and significant *Apc* disruption.

Whether or not it implies a loss of tumour suppressor function on the $Apc^{Min/+}$ background, the many statistically significant differences between WT and $Apc^{Min/+}$ mice, with regard to both transcript level and transcript distribution, suggests some level of interaction between the Wnt pathway in which Apc acts (Sansom et al., 2004) and the as-yet-unknown C11orf pathway. As the Wnt pathway, and Apc in particular, have strong established involvement with CRC (Powell et al., 1992), it is encouraging to see this evidence of a link with our CRC risk locus.

Chapter 5: Creation and Characterisation of Knockout Mouse Models

5.1 Introduction

Mouse models have long been established as a key tool to study biological systems. Their similarities to humans, on a genetic and physiological level, make them good representations of the human system. Of course, we must always remember they are models and do not completely represent human biology; in addition to the interspecies differences, the highly homogenous genetic background and closely regulated environments of laboratory mice are not representative of human populations. Nevertheless, mouse models have contributed to many important insights.

They are of particular use in GI research, as their digestive systems are far more similar to ours than other common models such as *C. elegans*, *Drosophila* or zebrafish. With respect to the 11q23.1 locus, only placental mammals have orthologues of all three genes, which of course limits the models that can be used. Additionally, as mice are an established CRC model, there are many research tools available, such as well-characterised colon-specific carcinogens and CRC-susceptible genetic backgrounds. While larger animal models can be a closer match in terms of their digestives systems, mice also have the benefit of their size and lifespan making them a more economical model than pigs or primates.

The equivalent of the 11q23.1 locus in mice contains orthologues of all three genes, with sequence conservation ranging 56-82%. As shown by my RNAscope results in chapter 4, the pattern of gene expression within colorectal tissue is consistent between human and mouse. I therefore created a knockout mouse model to provide information on the *C11orf* genes' functions by observing the effect of their loss on a complete organism. I used CRISPR/Cas9 to generate a 20kb deletion spanning the 3' end of *C11orf53*, all of *C11orf92* and the 5' end of *C11orf93*. I focused on studying this triple deletion, but future work on models with only a single gene knockout is planned.

I investigated the model's CRC risk both by allowing the mice to age, and by using genetic and chemical means of tumour induction. We also used various techniques to assess the mice, including blood counts, intestinal expression analysis, and immunohistochemistry (IHC) on intestinal tissue. Additional data came from observations on the breeding and survival of

the line. Since this is a novel model and the function of the genes is unknown, we took a broad approach to studying the animals, such as using a basic panel of intestinal markers for IHC. As we gather more data and improve our understanding of the genes' functions, we can carry out more targeted experiments to measure specific features of interest.

5.2 Generation of the *C11orf* mouse models

5.2.1 Methods

I used the CRISPR/Cas9 system to create mutations in C57BL/6J mouse embryos. This strain was chosen because the group's existing mouse models were also C57BL/6J and we wanted a consistent genetic background for any future crosses between lines. An injection mix of gRNAs and Cas9 mRNA was injected into embryos by Emma Allen at the transgenic service. No repair template was used, as I did not aim to create specific mutations using homology-directed repair, but to induce random indels via non-homologous end joining.

As it was not known to what extent, if at all, each of the three genes of the locus is contributing to CRC risk, we decided to begin by disrupting all three. Due to the close proximity of the genes we could not rely on recombination to generate a compound mutant from separately-created lines. I therefore included a gRNA for each gene in a single injection mix, with the aim of inducing an indel into each gene. We also hoped that, since CRISPR is not 100% efficient, I would also generate single and double gene knockout mutations alongside the triple gene knockout. If the triple gene knockout displayed an altered phenotype we could then proceed with analysing individual gene knockouts.

I designed gRNAs to target exonic regions of *C11orf53* and *C11orf93* approximately one third into the coding sequence (figure 5.1a), with the intention that this was far enough into the gene that it would not be skipped by a late alternative start site, but with enough sequence remaining that a frameshift mutation would have a significant impact. I also ensured the exons targeted were common to all transcripts. For *C11orf92*, this restricted the target region to a single small exon which forms the entirety of one transcript, but which is within a larger exonic region in the other transcript. gRNA sequences were designed using the online CRISPR design tool from the Zhang lab at MIT (www.crispr.mit.edu (Hsu et al., 2013)).

I designed genotyping primers approximately 400bp either side of the each gRNA target site, to allow sequencing across any indels while being far enough away to avoid disruption of the primer sites (figure 5.1a).

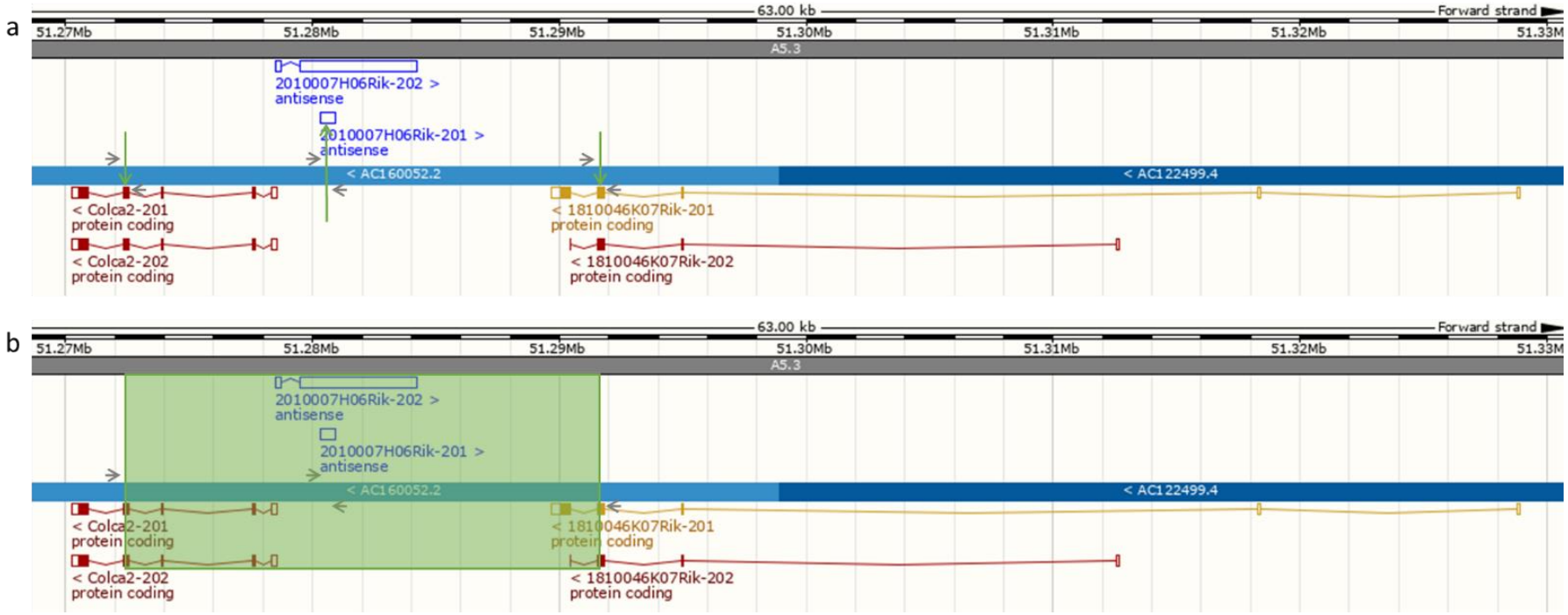


Figure 5.1: The region of mouse chromosome 9 containing the orthologues of *C11orf53*, *C11orf92* and *C11orf93*. (a) The sites targeted by gRNAs to generate knockouts by CRISPR are highlighted with green arrows. Positions of genotyping primers are illustrated with grey arrows (not to scale). (b) The *C11orf*$^{-/-}$ mouse line has a deletion covering 20kb, represented in green. Use of the outer two primers gives a product of 800bp when the deletion is present, while the inner two primers do not give a product. Images adapted from Ensembl, July 2018.

5.2.2 Results

Following the microinjection session and subsequent transfer of embryos into pseudo-pregnant females, 22 pups were born, of which 20 survived to 16 days for genotyping. I performed PCR using the three genotyping primer pairs. Only 11 of the mice had a product for all three genes, with four giving two products, three giving one product, and two mice having no PCR products for any primer pair. I used additional primers within GAPDH to confirm the DNA extractions had been successful.

I was able to complete genotyping of the mice by recombining primer pairs. The size of the products indicated large sections of sequence, spanning the distance between gRNA target sites, had been removed (figure 5.1b); this was confirmed by sequencing. Mutation results are summarised in table 5.1. While *C11orf53* and *C11orf92* show similar mutation frequencies, *C11orf93* was less frequently altered. This may mean that disruptions of this gene are less tolerated than the other two genes. Alternatively, it may be that the gRNA for this gene was not as efficient as the other two gRNAs, although the score assigned to each gRNA sequence by the Zhang lab online tool ranked the *C11orf93* gRNA above the *C11orf92* gRNA. As *C11orf92* is believed to be non-coding, it is much more difficult to predict if a small indel will have any impact on gene function, compared to the induction of a premature stop codon or a frameshift in a protein-coding gene; this is why the proportion of disruptive mutations in lower for this gene.

I chose three independent mutations to take forward and study. All three had 20kb of sequence lost, splicing exon 5 of *C11orf53* onto exon 4 of *C11orf93* and completely removing *C11orf92*, and as such totally disrupting all three genes. The other mutations were cryopreserved. In addition to sequencing, mRNA-seq expression analysis (discussed below) shows complete loss of transcripts for the three deleted genes in the homozygous knockout, with an intermediate level of expression in the heterozygote (figure 5.2a-5.2c). The exception is *C11orf93* in the proximal small intestine, which shows unaltered expression. As the downstream section of this gene is still present, spliced to the upstream section of *C11orf53*, it is possible that the splice gene is still being transcribed, particularly since *C11orf53* ordinarily has high expression in this tissue. There is not a corresponding peak for *C11orf53* in the knockout, but as only half of this gene is present it may not be detected by the alignment.

CRISPR carries with it the potential for off-target effects. An additional 5' G was included in the gRNAs to reduce these (Cho et al., 2014), but this does not guarantee their absence. We have not carried out whole genome sequencing to confirm that there are no such mutations. However, the three independent triple knockout lines all exhibit the same phenotype, and the phenotype continues to segregate with the *C11orf* mutation after six generations. I am therefore confident that the observed phenotype is due to the loss of the target locus. mRNA-seq expression analysis of the closest gene, *Pou2af1*, around 30kb away, shows no loss of expression and so the deletion is contained to the region of interest (figure 5.2d).

I subsequently re-derived a line carrying an 11bp deletion in *C11orf53*, which creates a premature stop codon after 20 amino acids, to examine the loss of *C11orf53* alone. This line is still in its early stages so only some data is available for this mutation. None of the mice resulting from my CRISPR experiment carried a lone *C11orf93* mutation, but a knockout line has been generated by the International Mouse Phenotyping Consortium (IMPC). Due to a husbandry error we have not as yet been able to establish the line in our unit for our own study but where applicable reference will be made to publically available IMPC data, which is summarised in figure 5.3.

Gene	Mutations	Disruptive mutations
C11orf53	32/40	29/32
C11orf92	31/40	21/31
C11orf93	11/40	10/11

Table 5.1: Frequency of mutations of each gene in the mice resulting from the initial CRISPR injection, based on 40 copies of each gene across 20 pups, and the number of those mutations that can be reasonably expected to disrupt the gene by introducing a frameshift mutation or an early stop codon.

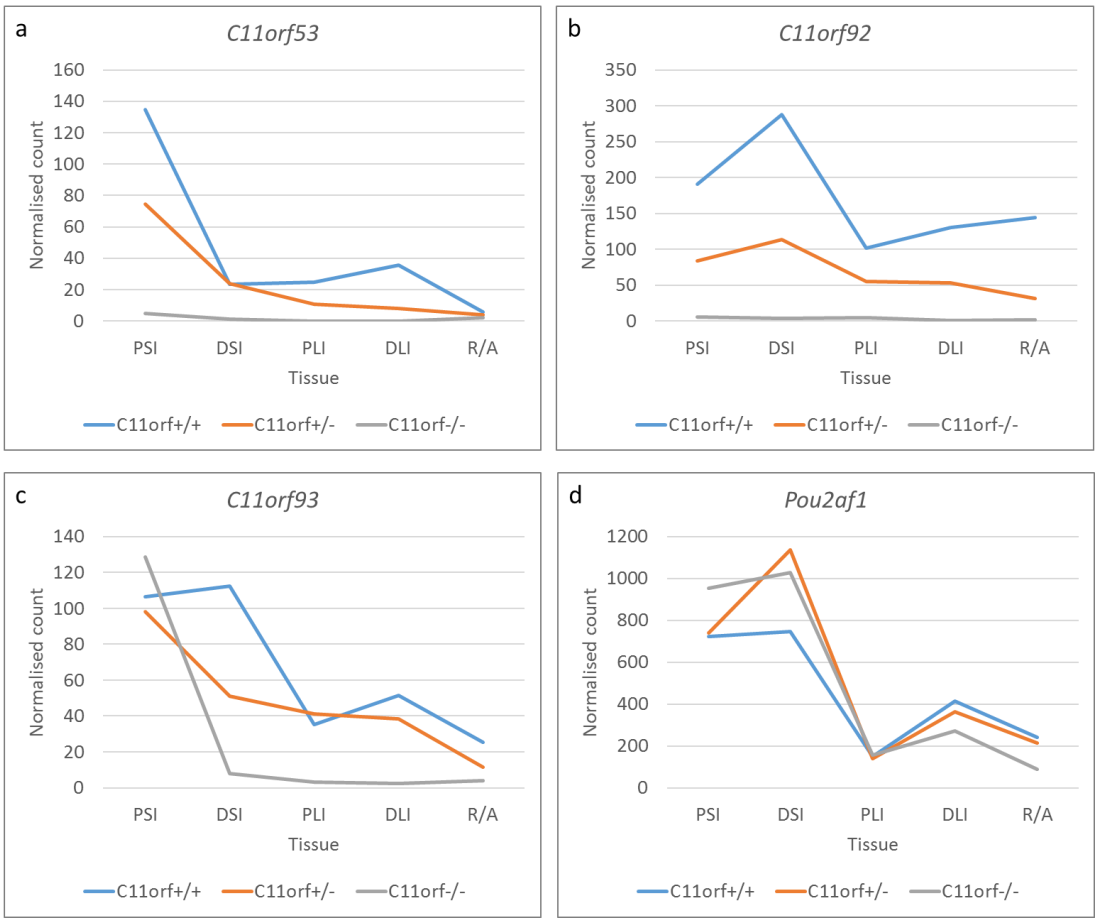


Figure 5.2: Normalised counts of mRNA-seq expression data across proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and the rectum/anus for (a) *C11orf53*, (b) *C11orf92*, (c) *C11orf93* and (d) *Pou2af1*, the next nearest gene to the deleted locus. The *C11orf*$^{-/-}$ mice show large reduction in their expression of *C11orf53*, *C11orf92* and *C11orf93* compared to *C11orf*$^{+/+}$, with an intermediate reduction in *C11orf*$^{+/-}$. *Pou2af1* expression is largely unchanged between the three genotypes.

Figure 5.3: All traits measured in the *C11orf93* knockout line generated by the IMPC, with the p-values from those measurements. Image taken from www.mousephenotype.org

5.3 Maintenance of the lines

5.3.1 Methods

Mice were housed in individually ventilated cages (IVCs) with group sizes and enrichment according to Home Office regulations. The cage room was kept at 21-23°C with a 12-hours-on, 12-hours-off light cycle all year. The unit has a pathogen health status rating of 4. Mice were given the RM3 diet (Special Diets Services) in pellet form, a high nutrient diet containing fatty acids, with constant access to both food and water. Pups were earclipped for genotyping at 14 days old and removed from their mother into single sex caging at 21 days old.

5.3.2 Results

5.3.2.1 Breeding

Male *C11orf*$^{-/-}$ mice are infertile. This was demonstrated by caging two *C11orf*$^{-/-}$ males with three females each, all animals being of breeding age, for three months with no pups resulting. The IMPC lists infertility in male *C11orf93* homozygous knockouts. No fertility issues have been observed in females of any genotype.

Genotype frequencies in the *C11orf* knockout line do not fit what would be expected from Medelian ratios. In the pups resulting from *C11orf*$^{+/-}$ x *C11orf*$^{+/-}$ pairings, there is a significant decrease in the incidence of homozygote mutants, implying a developmental issue resulting from being *C11orf* null that reduces viability (table 5.2a). There is also a significant increase in heterozygote incidence, suggesting some form of heterozygote advantage. *C11orf*$^{+/+}$ x *C11orf*$^{+/-}$ matings show an increase in heterozygote incidence from expected, although it is not significant (table 5.2b). *C11orf*$^{+/-}$ x *C11orf*$^{-/-}$ matings show a significant decrease in homozygote incidence (table 5.2c).

By comparison, the *C11orf53* mutant line do not show deviations from the expected genotype frequencies (table 5.2d). IMPC data shows the genotype ratios in the *C11orf93* line are as expected. No differences in sex distribution have been observed in any genotypes.

As part of our tumour induction studies, discussed below, I crossed the *C11orf* knockout line onto the *Apc*^{Min} background. This mouse strain carries a homozygous-lethal nonsense

mutation in the Wnt antagonist *Apc*, causing constitutive activation of the Wnt signalling pathway and leading to spontaneous intestinal adenoma formation. Crossing onto this background can be used to accelerate tumourigenesis (Su et al., 1992). The differences between the expected and observed ratios of genotypes becomes more pronounced when at least one parent carries the Apc^{Min} mutation, compared to both parents being *Apc* wild type (table 5.3). While the $C11orf^{-/-}$ and $Apc^{Min/+}$ mutations are not entirely incompatible – I have had one $C11orf^{-/-}$ $Apc^{Min/+}$ mouse – these results do suggest an interaction between the two pathways that reduces viability beyond the two mutations alone.

To investigate at what point the homozygotes are being lost, I collected embryos at E14.5 from $C11orf^{+/-}$ $Apc^{+/+}$ females mated with $C11orf^{+/-}$ $Apc^{Min/+}$ males. Embryos were fixed and whole mounted, with the yolk sacs being used for genotyping. There was no significant deviation in genotype frequencies from expected (table 5.4). All embryos were still alive at the point of collection. Histology does not show any developmental defects in any of the embryos and there was no difference in size between the genotypes.

As embryos at E14.5 are too big to be reabsorbed, all homozygotes still present after this point are being born, but we do not know if they are alive at birth. The litter size, as counted by the technicians, does not change, meaning all pups present when the litter is first counted are still alive when genotyping samples are taken at day 14; however, there can be several hours between a litter being born and being counted during which a mother can eat pups. Recording the cages with a night-vision camera could be used to determine if the pups are dying in the last days of gestation and being born dead or are dying/being killed shortly after birth.

5.2a) *C11orf*$^{+/-}$ x *C11orf*$^{+/-}$ matings – 261 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/+}$	1	65.25	1	67
C11orf$^{+/-}$	2	130.5	2.46	165****
C11orf$^{-/-}$	1	65.25	0.43	29****
5.2b) *C11orf*$^{+/+}$ x *C11orf*$^{+/-}$ matings – 28 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/+}$	1	14	1	11
C11orf$^{+/-}$	1	14	1.55	17
5.2c) *C11orf*$^{+/-}$ x *C11orf*$^{-/-}$ matings – 15 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/-}$	1	7.5	1	12*
C11orf$^{-/-}$	1	7.5	0.25	3*
5.2d) *C11orf53*$^{+/-}$ x *C11orf53*$^{+/-}$ matings – 25 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf53$^{+/+}$	1	6.25	1	7
C11orf53$^{+/-}$	2	12.5	1.71	12
C11orf53$^{-/-}$	1	6.25	0.86	6

Table 5.2: Expected and observed frequencies of pups resulting from matings between a) two *C11orf* heterozygous knockouts, b) *C11orf* wild type with *C11orf* heterozygous knockout, c) *C11orf* heterozygous knockout with *C11orf* homozygous knockout, and d) two *C11orf53* heterozygous knockouts. Significant deviations from expected, according to chi squared testing, have been marked. Table 5.2a contains both *Apc*$^{+/+}$ and *Apc*$^{Min/+}$ mice; this data is split in table 5.3a and 5.3b. Tables 5.2b, 5.2c and 5.2d contain only *Apc*$^{+/+}$ mice.

5.3a) *C11orf$^{+/-}$Apc$^{+/+}$* x *C11orf$^{+/-}$Apc$^{+/+}$* matings – 184 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/+}$	1	46	1	49
C11orf$^{+/-}$	2	92	2.2	108*
C11orf$^{-/-}$	1	46	0.55	27**
5.3b) *C11orf$^{+/-}$Apc$^{Min/+}$* x *C11orf$^{+/-}$Apc$^{+/+ \text{ or } Min/+}$* matings – 77 pups				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/+}$	1	19.25	1	18
C11orf$^{+/-}$	2	38.5	3.17	57****
C11orf$^{-/-}$	1	19.25	0.11	2****

Table 5.3: Expected and observed frequencies of pups resulting from matings between two *C11orf* heterozygous knockouts, as given in table 5.2a, divided between matings where (a) both parents were *Apc$^{+/+}$* and (b) at least one parent was *Apc$^{Min/+}$*. Significant deviations from expected, according to chi squared testing, have been marked.

5.4) *C11orf$^{+/-}$Apc$^{+/+}$* x *C11orf$^{+/-}$Apc$^{Min/+}$* matings – 20 embryos				
Genotype	**Expected**		**Observed**	
	Ratio	**Number**	**Ratio**	**Number**
C11orf$^{+/+}$	1	5	1	4
C11orf$^{+/-}$	2	10	3.25	13
C11orf$^{-/-}$	1	5	0.57	3

Table 5.4: Expected and observed frequencies of embryos collected at E14.5 resulting from matings between two *C11orf* heterozygous knockouts where the mother was *Apc$^{+/+}$* and the father was *Apc$^{Min/+}$*.

5.3.2.2 Survival

In addition to their reduced incidence at birth, the *C11orf*$^{-/-}$ mice display a striking failure to thrive, with greatly reduced survival by adulthood. All homozygous knockout mice survive to 20 days of age, but by day 31 only 52% are still alive (figure 5.4). Those mice that die typically exhibit a runt phenotype. Those that do survive past one month old display no health problems past one year old, although on average they are smaller than their littermates (figure 5.5). Heterozygous knockouts do not show any survival problems. In comparison, no survival problems have been observed in the *C11orf53* mutant line or are listed in the IMPC *C11orf93* knockout line data, although the *C11orf93* knockout mice do show significantly reduced body weight.

The timing of this extreme drop in survival is highly interesting, as the mice are weaned at 21 days. Although pups may begin to eat some solid food before this point, at 21 days they are removed from their mother to new cages, marking a complete transition to purely solid food with no milk. Access to soft mash in addition to hard pellets after weaning did not aid survival. There is no pattern of survival correlating to litter size, to maternal age, to maternal genotype, or whether it was the mother's first litter.

As the deaths appeared to be triggered by weaning, we tried delaying weaning from 21 days to 28 days. This did not improve survival, which actually decreased for this cohort to 20%. There was higher proportion of animals that were found partially eaten or missing entirely, compared with the 21 day cohort which were mostly found intact. It may be that some of the 28 day group, while lagging behind their littermates in terms of size, would not have died naturally, but were instead killed by their mothers. Alternatively, it may simply be due to the low numbers of animals in this group. Only three litters, containing five homozygous knockouts, were weaned late, as once the worsened survival was observed the protocol was ended and all litters returned to 21 day weaning.

The combination of reduced frequency and poor survival has severely limited the number of *C11orf*$^{-/-}$ mice that have been available to use. Much of the subsequent data in this chapter is based on low numbers of animals for this reason.

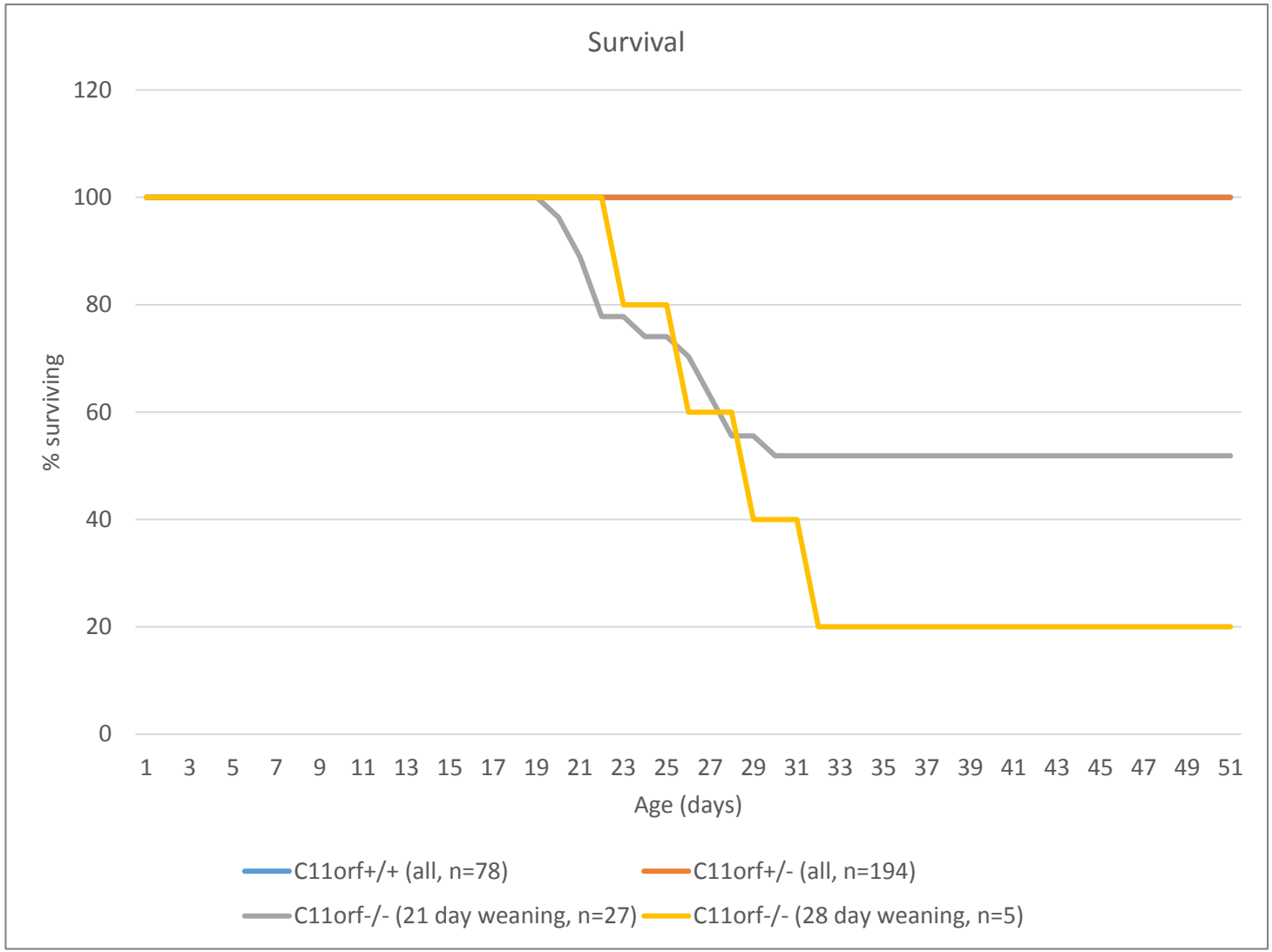


Figure 5.4: Survival curve for the *C11orf* knockout colony. *C11orf*$^{+/+}$ and *C11orf*$^{+/-}$ mice have 100% survival (the wild type line being directly under the heterozygous line, and so not visible), while *C11orf*$^{-/-}$ mice have greatly reduced survival at both 21 day and 28 day weaning. No differences were observed in *C11orf*$^{+/+}$ or *C11orf*$^{+/-}$ mice between 21 day and 28 day weaning and so these cohorts have been combined.

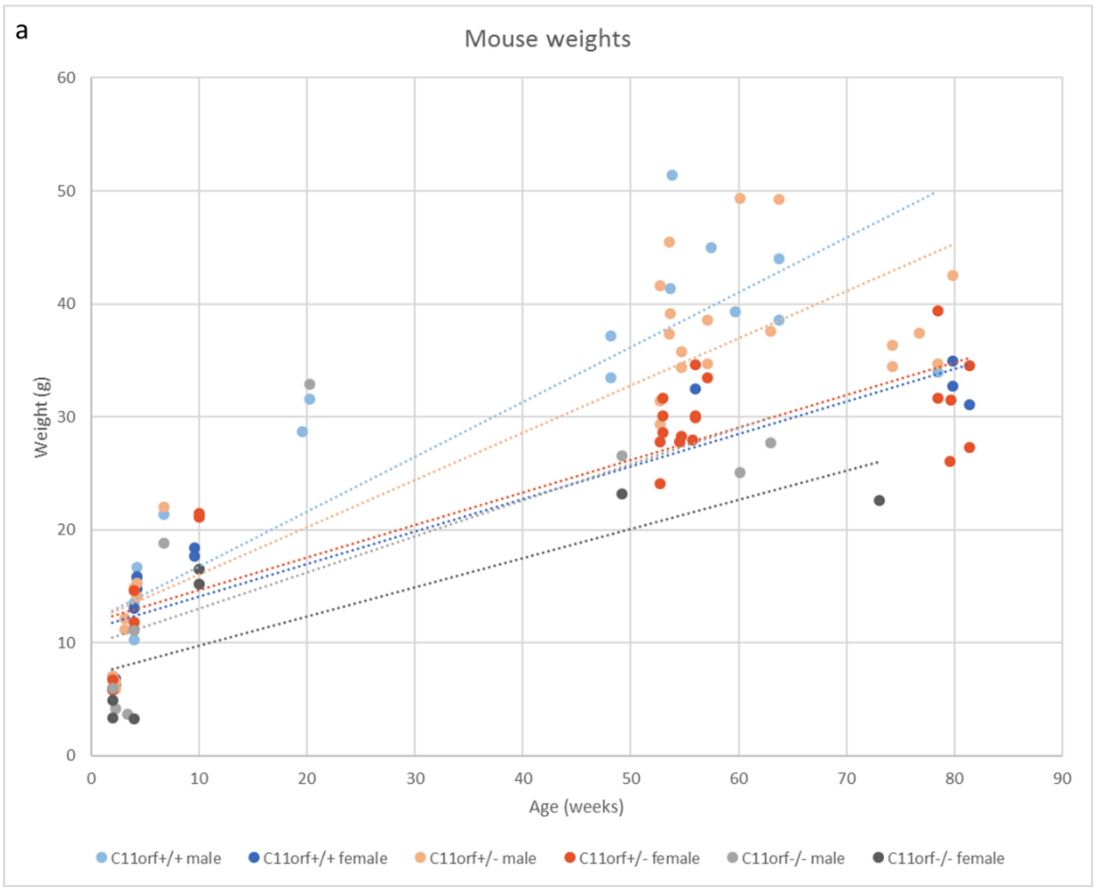


Figure 5.5a: Weights of *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice at various age points, coloured by genotype and sex.

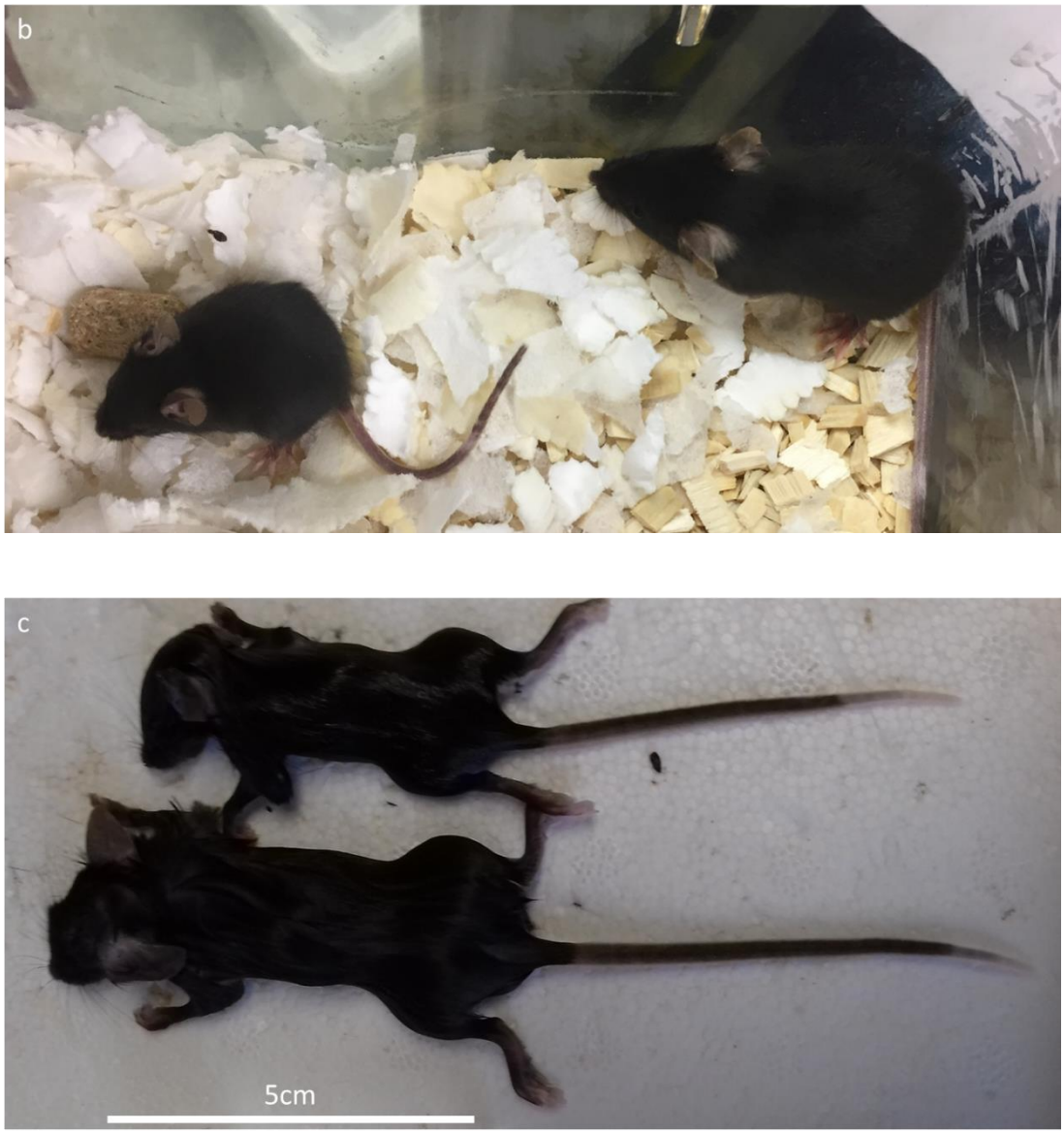


Figure 5.5: (b) Photograph of four-week old females, showing size difference between *C11orf*$^{-/-}$ (left) and *C11orf*$^{+/-}$ (right). Photo taken by Vidya Rajasekaran. (c) Photograph of three-week old males, showing size difference between *C11orf*$^{-/-}$ (top) and *C11orf*$^{+/+}$ (bottom).

5.4 Tumour induction

5.4.1 Methods

5.4.1.1 Aging

I aged the mice beyond one year old to allow any potential age-induced phenotypes to develop. Mice were monitored for signs of general and intestinal discomfort, including lethargy, hunched gait, rectal prolapse, blood around the anus, and pale extremities (an indication of anaemia). I collected intestines, caecum, stomach, liver, kidneys, spleen, heart and lungs, which were fixed in 10% NBF and paraffin embedded.

5.4.1.2 Chemical induction

1,2-dimethylhydrazine (1,2-DMH) is a potent alkylating agent which, when administered by subcutaneous injection, induces colonic tumours in mice and rats by causing G→A transitions (Fiala, 1977; Thumherr et al., 1973), with approximately two thirds of tumours having mutations in the *K-ras* proto-oncogene (Jacoby et al., 1991).

10 week old females were given subcutaneous injections of 40mg 1,2-DMH/kg body weight. The DMH was diluted in EDTA and we included mice injected with EDTA only as controls. We culled the mice after one week and harvested tissue. I collected intestines, caecum, stomach, liver, kidneys, spleen, heart and lungs, which were fixed in 10% NBF and paraffin embedded.

5.4.1.3 Susceptible genetic background

The Apc^{Min} mouse line, as previously discussed, is a model of familial adenomatous polyposis. As it has a very reliable development of intestinal polyps, it is frequently used to assess the effects of other potential CRC risk factors which alone may not be sufficient to induce CRC within the time constraints of a mouse model, but which can amplify the $Apc^{Min/+}$ phenotype.

I crossed the *C11orf* knockout line with the Apc^{Min} line. As described above, the incidence of $C11orf^{-/-}$ mice was extremely low on the Apc^{Min} background (table 5.4) and only one $C11orf^{-/-}$ $Apc^{Min/+}$ mouse was produced.

Mice were monitored for symptoms of discomfort, as in the aging cohort, and culled accordingly; the *Apc*$^{Min/+}$ phenotype is fatal if left to develop fully, but this was not permitted to happen and mice were culled when they began showing signs of distress. I collected intestines, caecum, stomach, liver, kidneys, spleen, heart and lungs, which were fixed in 10% NBF and paraffin embedded. Between fixation and embedding the intestines were stained with methylene blue, a dye which enters the crypts of the intestines. When the tissue is subsequently destained in ethanol, the dye remains in aberrant crypt foci, accentuating polyps and so aiding counting (Bird, 1987).

5.4.2 Results

5.4.2.1 Aging

Two male homozygous knockouts were culled at 14 months of age, as they had pale extremities. This is a sign of anaemia, and can indicate the presence of growths in the intestines restricting nutrient absorption. However, no sign of any growths was present when I examined the tissue, with all organs appearing normal (figure 5.6-5.12). I also collected blood for full blood counts, which did show anaemia with low red blood cells, and low white blood cells (see below for full blood analysis). One female homozygous knockout has reached 15 months of age without pale extremities.

Heterozygotes were aged to 18 months without exhibiting any symptoms of health issues, and histological examination does not reveal the development of any phenotype with age.

In total 11 male and eight female wild types, 27 male and 26 female heterozygotes, and three male and two female homozygotes have been aged to one year or beyond.

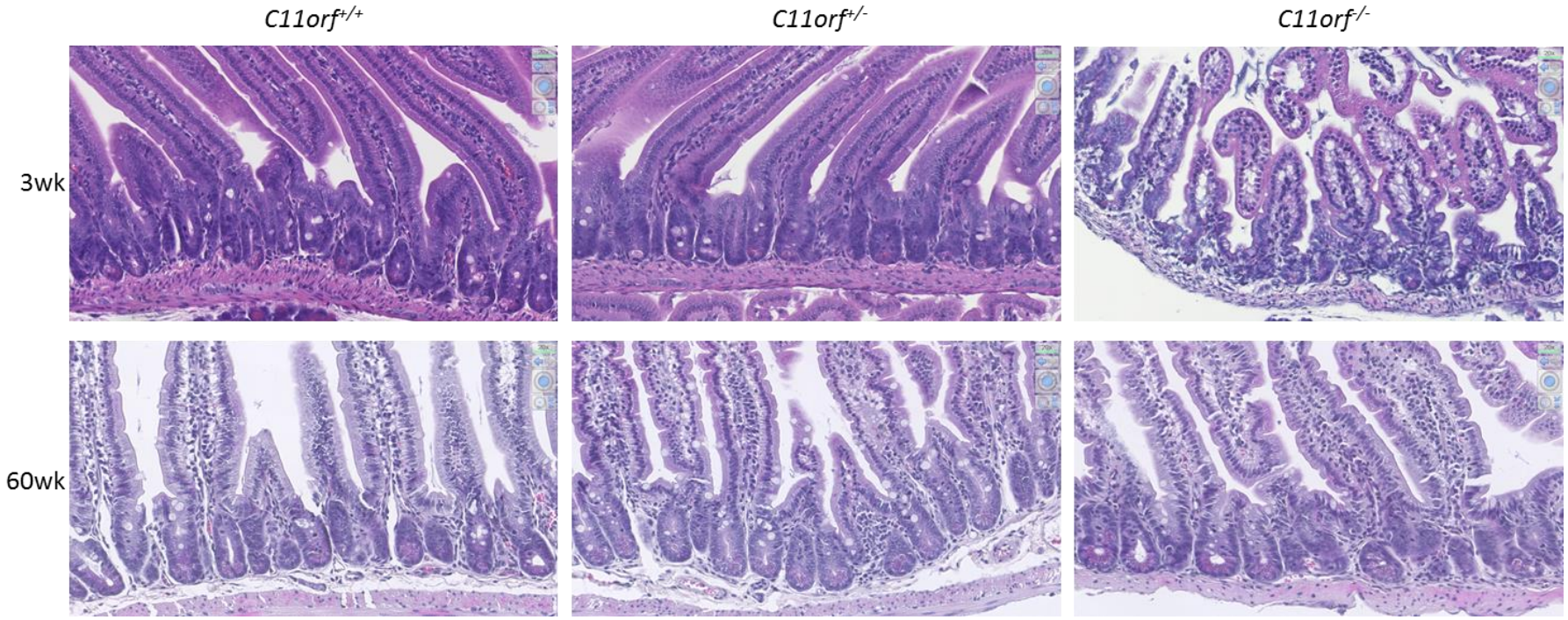


Figure 5.6: H&E sections of proximal small intestine collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

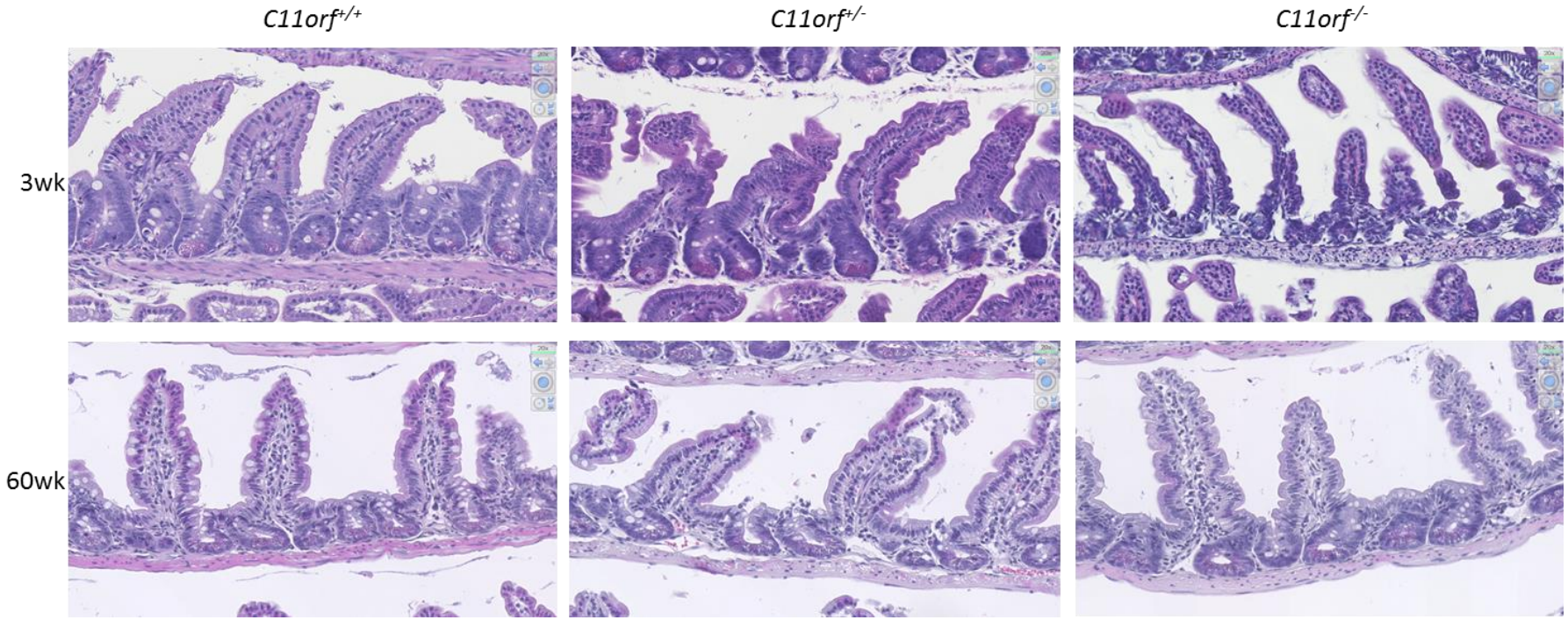


Figure 5.7: H&E sections of distal small intestine collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

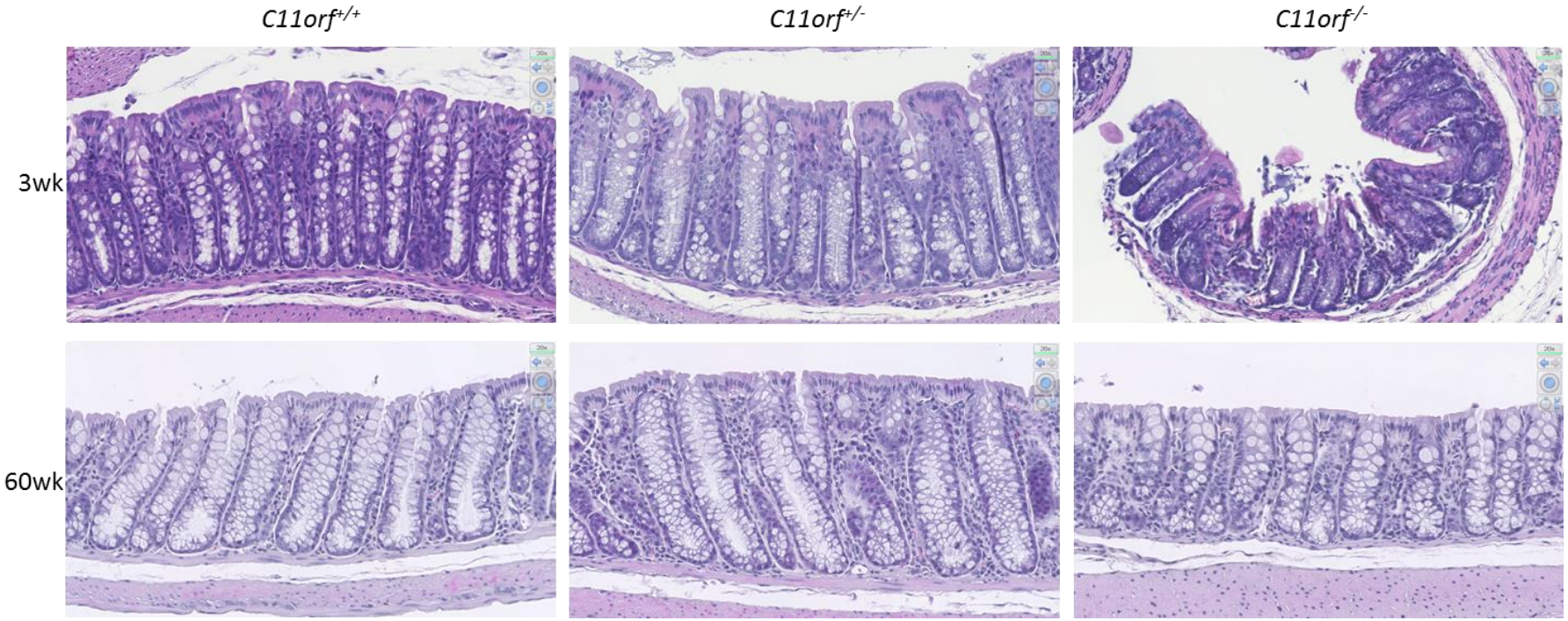


Figure 5.8: H&E sections of large intestine collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

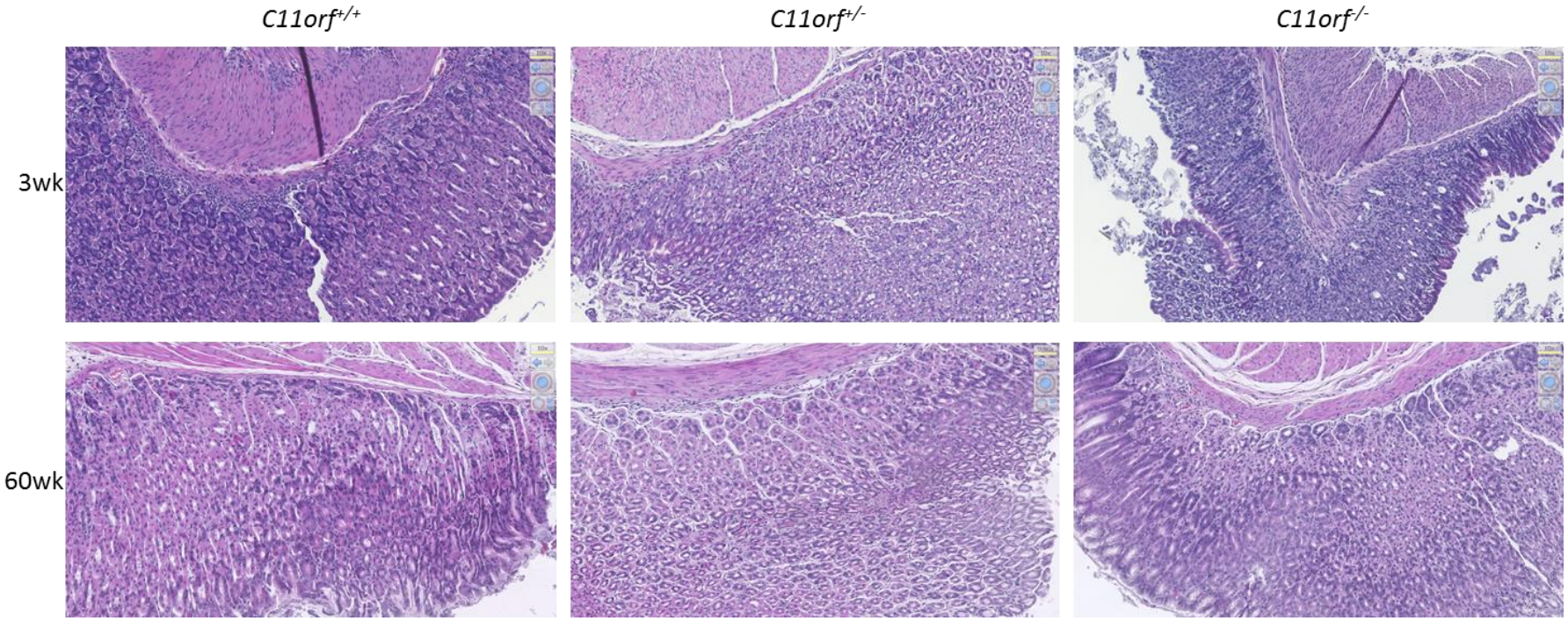


Figure 5.9: H&E sections of stomach collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 100X.

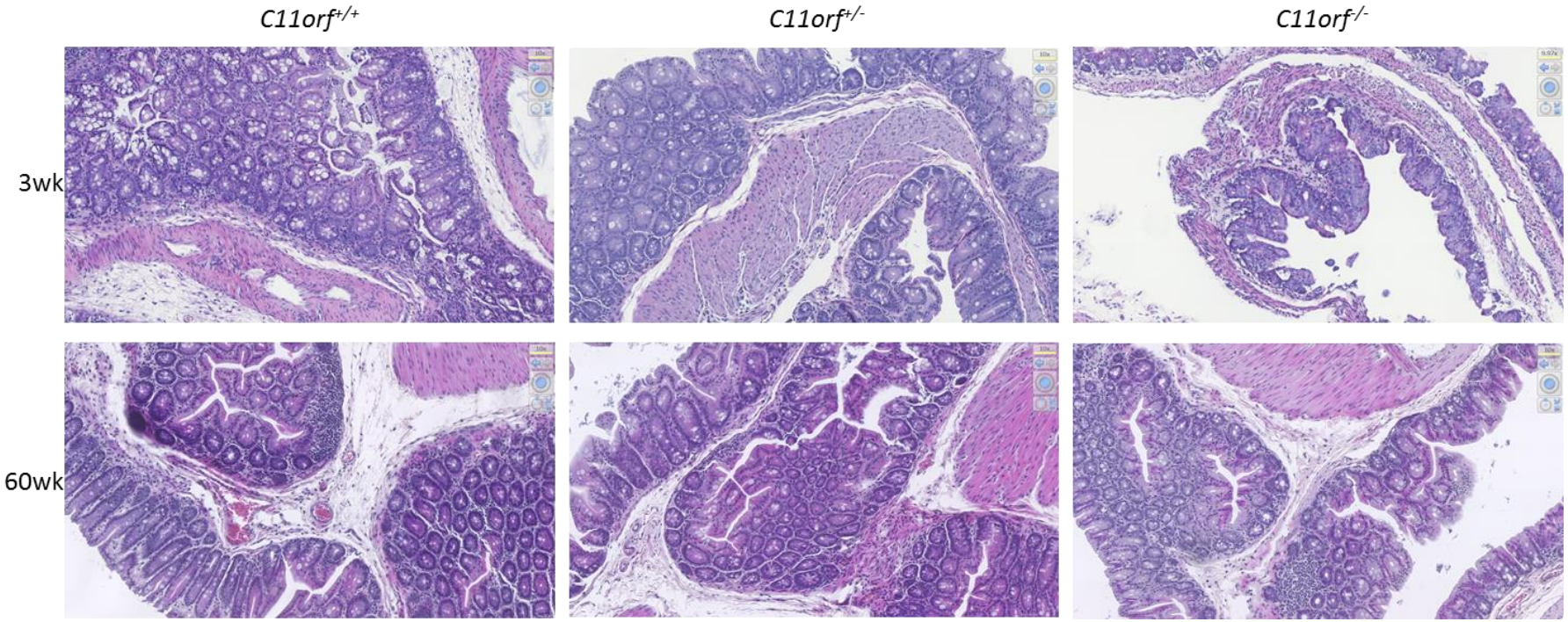


Figure 5.10: H&E sections of caecum collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 100X.

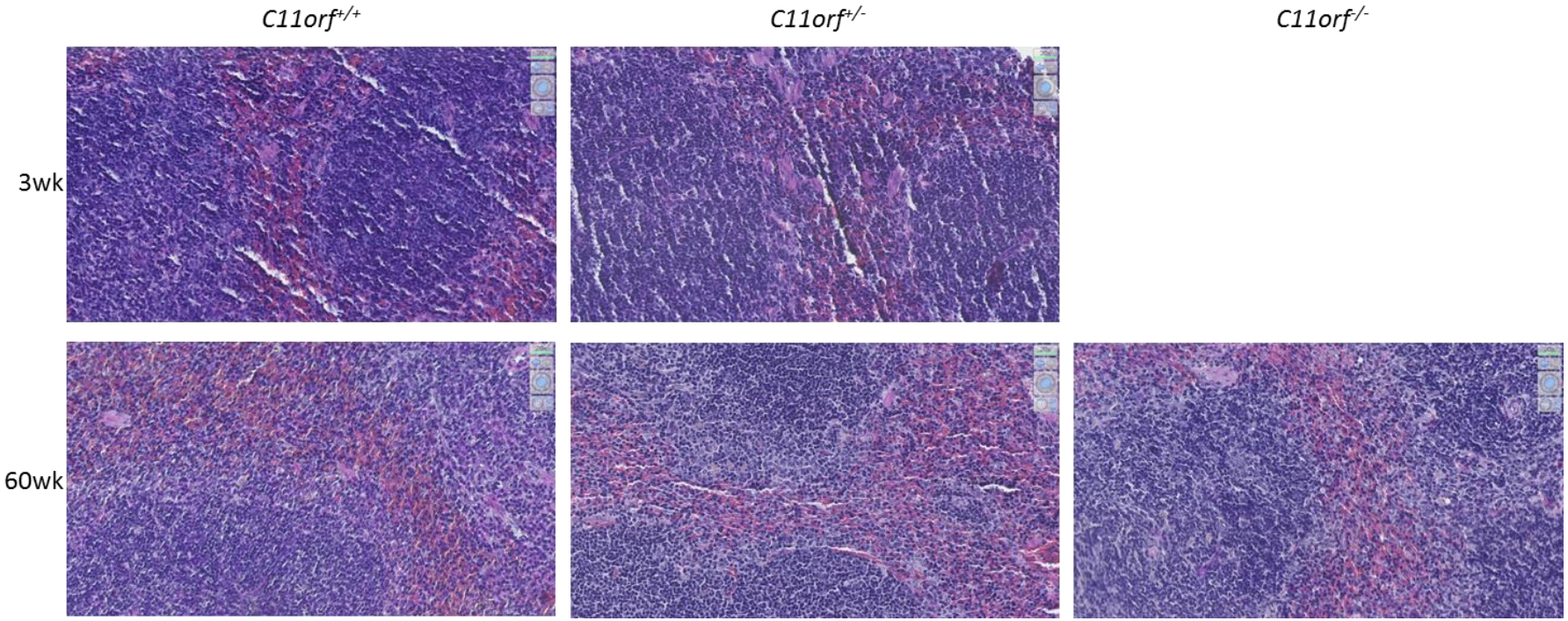


Figure 5.11: H&E sections of spleen collected at 3 weeks and 60 weeks of age, from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Due to the small size of *C11orf*$^{-/-}$ mice, the spleen was not identified at 3 weeks. Images taken at 200X.

5.4.2.2 Chemical induction

We saw no tumour development in mice of any *C11orf* genotype. While one week is not a long period of time, 40mg/kg is a high dose of 1,2-DMH and one week has been sufficient to induce a large growth in another of our mouse models when given 40mg/kg 1,2-DMH. There are multiple alternative regimes for DMH-induction, some of which we would have performed had we had the available mice, such as weekly injections of 20mg/kg 1,2-DMH for five weeks with tissue collection five weeks after the final injection, or 15mg/kg injections weekly injections for 12 weeks.

A further experiment which I was unable to do due to limitations of animals and time, but which would be a highly interesting future experiment, is to combine 1,2-DMH or its metabolite azoxymethane (AOM) with dextran sulphate sodium (DSS). When given in drinking water, DSS induces chronic inflammation, a key risk factor in CRC development (Kulaylat and Dayton, 2010). A study comparing the susceptibility of two mice strains to colitis-associated CRC, through the use of a single AOM injection followed by intermittent DSS administration, detected a CA-CRC susceptibility locus on mouse chromosome 9 (Van Der Kraak et al., 2010). While the locus, *Css4*, is very large at 41Mb, it does cover *C11orf53*, *C11orf92* and *C11orf93*; *Css4* is syntenic to regions on human chromosomes 11 and 15, including 11q23. We would therefore have liked to repeat the experiment performed in this study, in which a 10mg/kg AOM injection was followed by three 4-day periods of 3% DSS in drinking water, beginning one week after the AOM injection and with 17 days between DSS periods, and tissue was collected a various timepoints up to 18 weeks after the AOM injection.

5.4.2.3 Susceptible genetic background

There is a slight trend for heterozygous knockouts to have more SI and LI polyps than wild type mice, but it is not significant (figure 5.12a, 5.12b). There was no difference in the mean age at which the mice were culled (figure 5.12c). The single homozygote knockout had more LI polyps than the averages for both wild types and heterozygous knockouts, but this was still within the observed range for both other genotypes. This was also the youngest mouse to be culled, at 76 days; only four heterozygotes and one wild type were culled under 100 days.

Apc$^{Min/+}$ mice exhibit splenomegaly, due to high splenic haematopoiesis which is thought to be a response to impeded differentiation of haematopoietic tissues (You et al., 2006). *Apc*$^{Min/+}$ mice have lost the quiescent haematopoietic stem cell population in bone marrow (Lane et al., 2010) and Wnt signalling has been shown to regulate haematopoiesis (Luis et al., 2011). Due to this extramedullary haematopoiesis *Apc*$^{Min/+}$ mice are not anaemic despite the disruption in their bone marrow (You et al., 2006). I measured the length and width of spleens removed from *C11orf*$^{+/+}$ and *C11orf*$^{+/-}$ *Apc*$^{Min/+}$ mice when they were culled for *Apc*$^{Min/+}$ phenotype. While the *C11orf*$^{+/-}$ mice still had enlarged spleens compared to *Apc*^{WT} mice, they were significantly smaller than *C11orf*$^{+/+}$ (figure 5.12d).

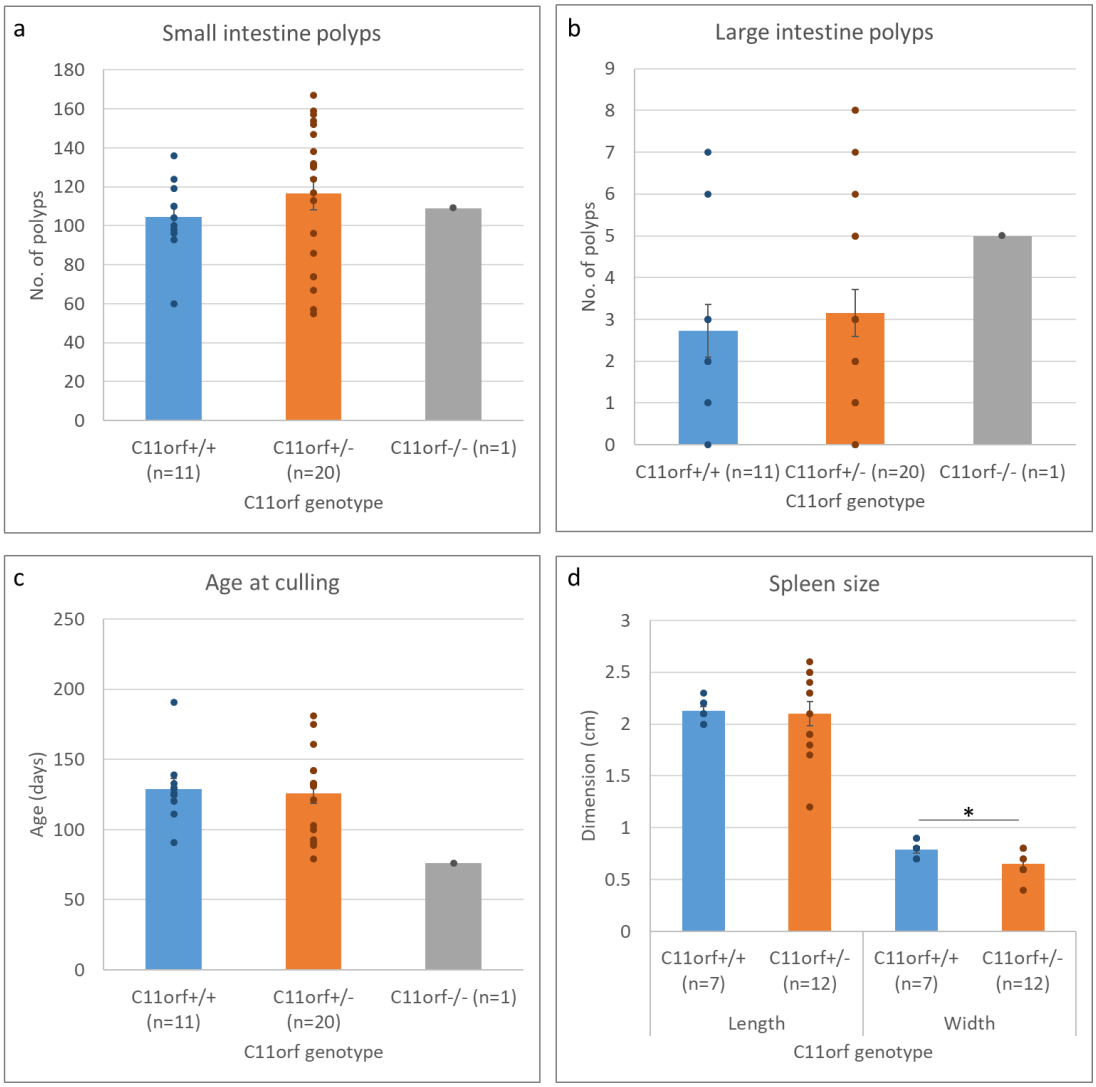


Figure 5.12: Phenotypes of *C11orf* mice on the *Apc*$^{Min/+}$ background. Bars represent the mean, with each animal shown as an individual datapoint. (a) Polyp counts in small intestine and (b) polyp counts in large intestine; (c) the age at which mice were culled due to *Apc*$^{Min/+}$ phenotype; (d) the size of their spleens at death. Data was not normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test. Significant results are indicated by stars.

5.5 Immunohistochemical analysis

5.5.1 Methods

We utilised immunohistochemistry (IHC) to visualise various proteins of interest in FFPE intestinal tissue from mice in the *C11orf* knockout line. IHC was performed by Vidya Rajasekaran, Marion Bacou and Calum Robertson on tissue from three-week old and 11-week old female mice, staining with antibodies raised against Ki-67, β-catenin, cytokeratin-7 and Clca1 and using DAB to visualise. Tissue staining was also carried out by the Pathology Department using periodic acid Schiff (PAS) and alcian blue for mucins. Images presented are representative of the wider tissue.

5.5.2 Results

Ki67 is expressed in proliferating cells, being present in S, G_2 and M phases but absent in G_0, with variable expression in G_1 (Gerdes et al., 1984). In healthy intestine, Ki-67 should only be present in the proliferating cells in the base of crypts. This is what is observed in the tissue of *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice, indicating there is no aberrant cell proliferation taking place outside the stem cell niche (figure 5.13).

β-catenin is a transcription factor which stimulates transcription of Wnt-responsive genes (Cong et al., 2003). It is present at low levels in the cytoplasm until Wnt signalling is activated, which inhibits β-catenin degradation, allowing it to accumulate and translocate into the nucleus (figure 1.4); high β-catenin levels in the nucleus is associated with poor prognosis in CRC (Chen et al., 2013b). In all *C11orf* genotypes, β-catenin is localised to the cytoplasm rather than the nuclei (figure 5.14).

Accumulation of nuclear β-catenin causes epithelial cells to lose cell-cell adhesion and cell polarity, and enter a dedifferentiated state (Mariadason et al., 2001; Naishiro et al., 2001). This epithelial-mesenchymal transition (EMT)-like change allows cancer cells to migrate and invade and so is key to progression (Brabletz et al., 2005). Cytokeratin-7, also called keratin-7 or sarcolectin, is a type II keratin expressed in epithelial cells and can therefore be used to detect regions where EMT has taken place (Ramaekers et al., 1990). No such loss of staining is observed in the *C11orf* knockouts (figure 5.15). Conversely, there appears to be stronger staining, which may mean that the cells are not maturing at the same rate as the other genotypes.

To detect mucins, the primary components of mucus, we used PAS and alcian blue staining. PAS detects neutral, acid-simple non-sulphated and acid-complex sulphated mucins, while alcian blue detects acid-simple non-sulphated and acid-complex mesenchymal mucins (Pernick, 2005). Both stains show a reduction in the knockout mice indicting a reduction in mucins (figure 5.16, 5.17).

Following the mucin staining results, we used an antibody for the goblet cell marker Clca1. Clca1 is a chloride ion channel regulator secreted at high levels in intestinal mucus to promote mucus expansion and aid correct mucus structure (Nyström et al., 2018). It is downregulated in CRC (Bustin et al., 2001; Yang et al., 2013). There is reduced staining between *C11orf*$^{+/+}$ and *C11orf*$^{+/-}$ mice, with a further reduction in *C11orf*$^{-/-}$ (figure 5.18). This loss of Clca1 is particularly pronounced in the base of the crypts.

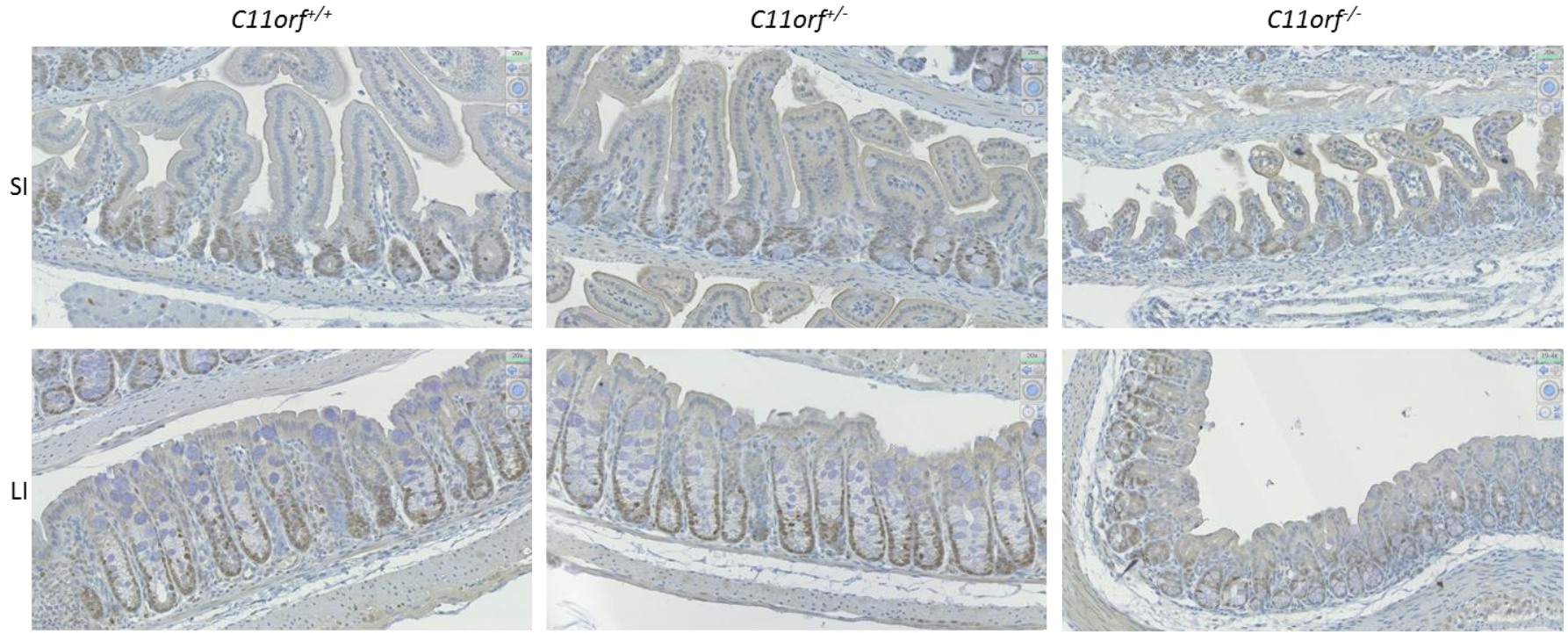


Figure 5.13: FFPE sections of intestinal tissue stained for proliferation marker Ki-67. Staining is only present in the base of the crypts, where the stem cell niche is located. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

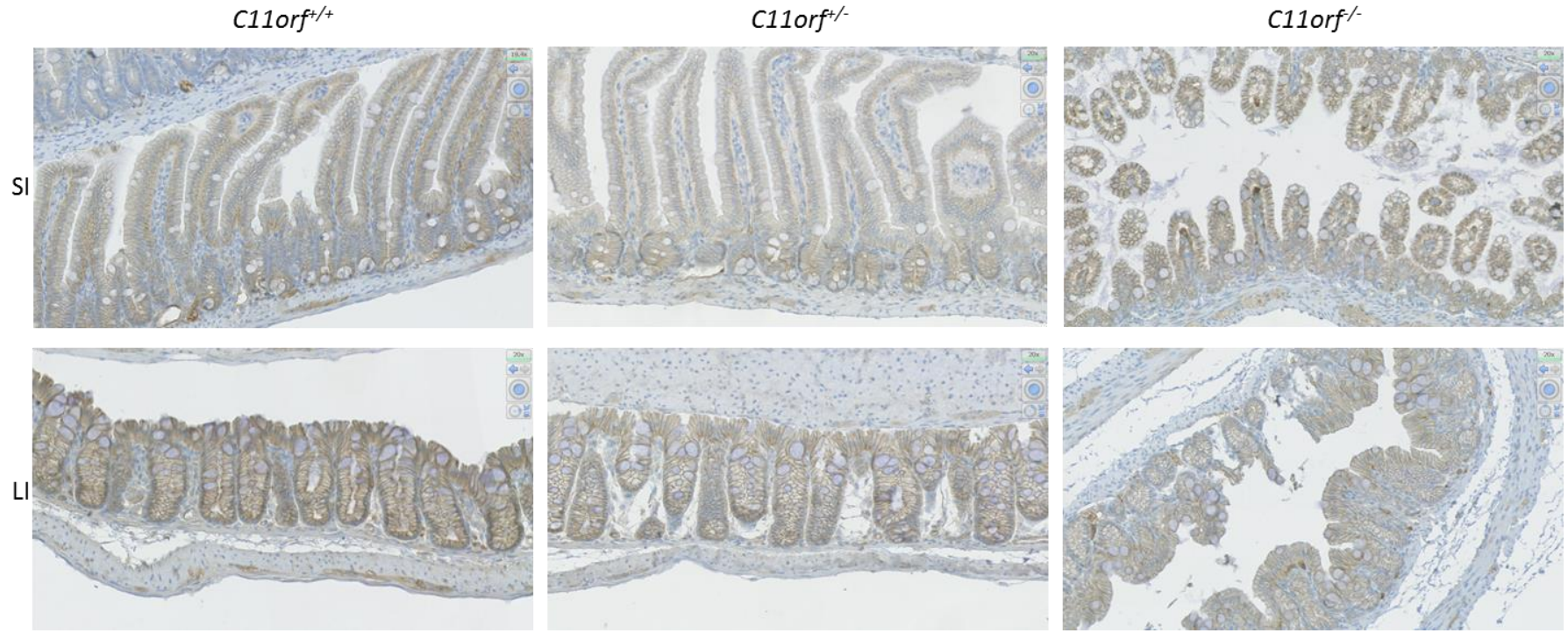


Figure 5.14: FFPE sections of intestinal tissue stained for transcription factor β-catenin. Staining is present in the cytoplasm of the cells but not in the nuclei. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

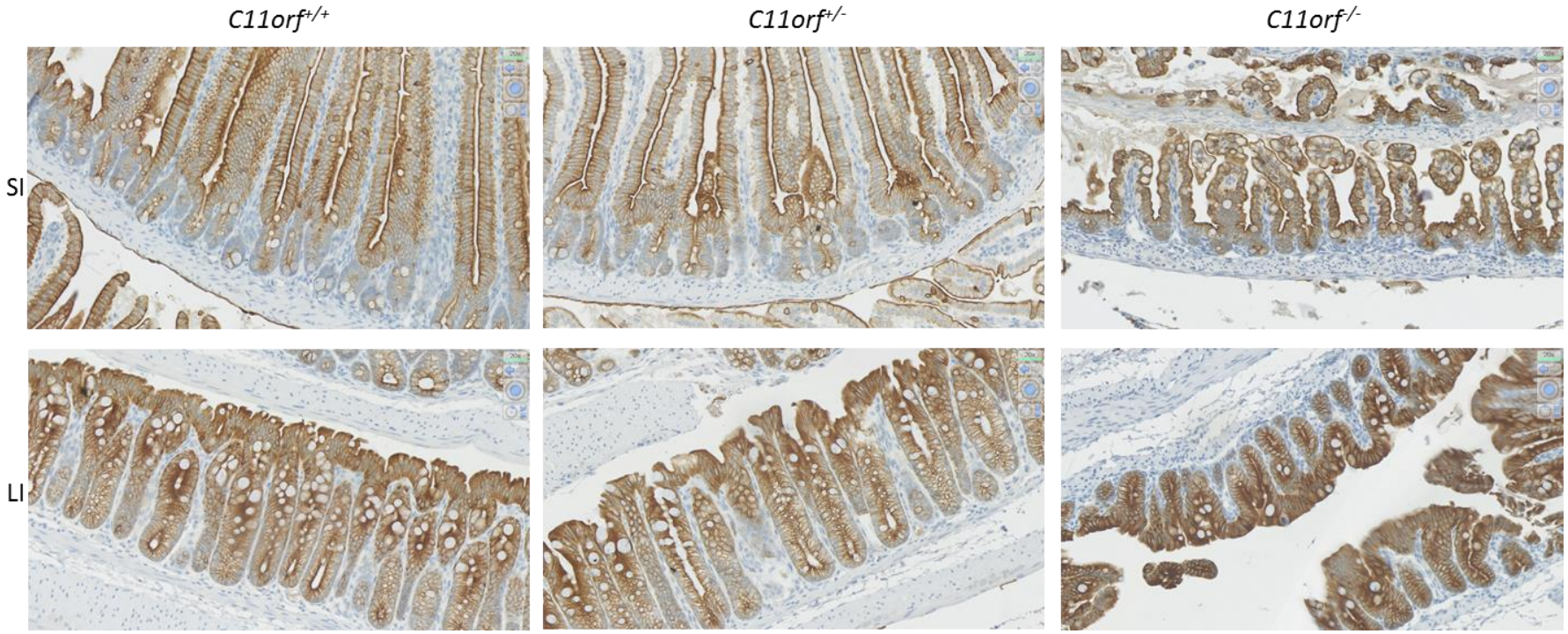


Figure 5.15: FFPE sections of intestinal tissue stained for epithelial marker cytokeratin-7. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

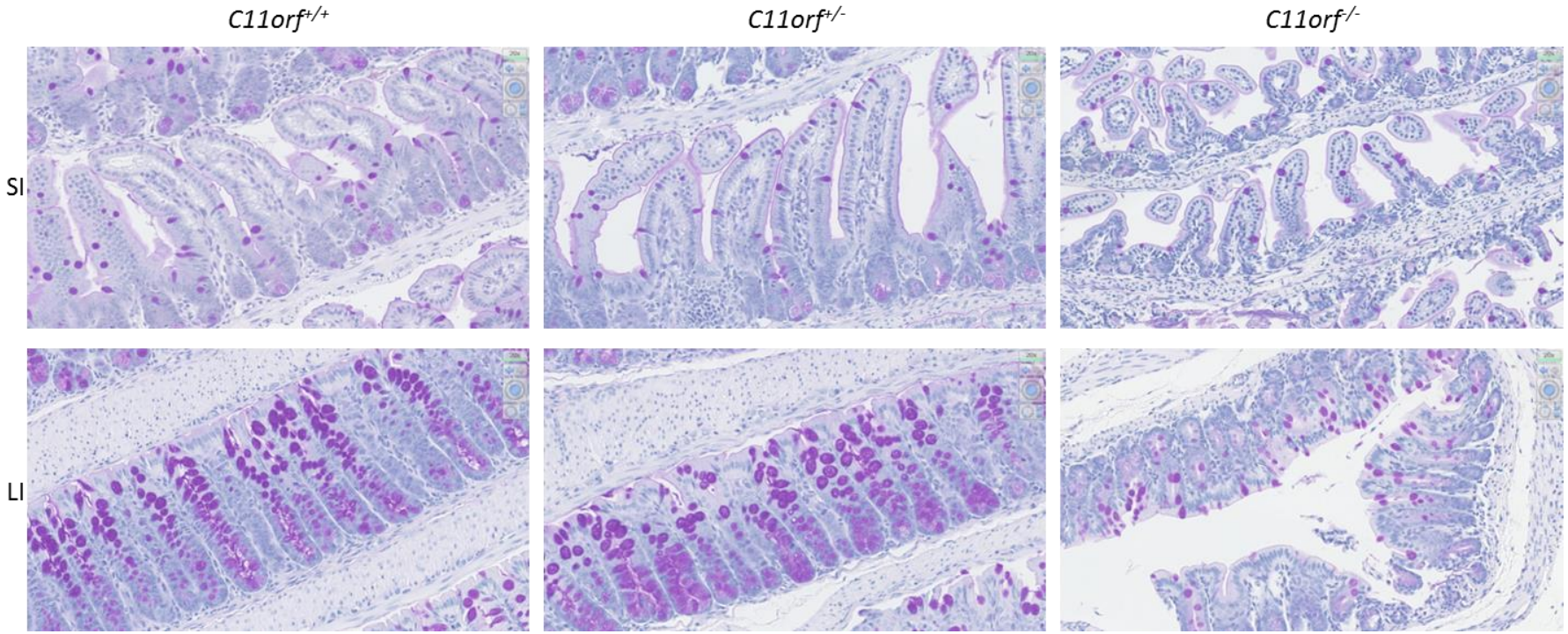


Figure 5.16: FFPE sections of intestinal tissue stained with periodic acid Schiff to detect neutral, acid-simple non-sulphated and acid-complex sulphated mucins. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

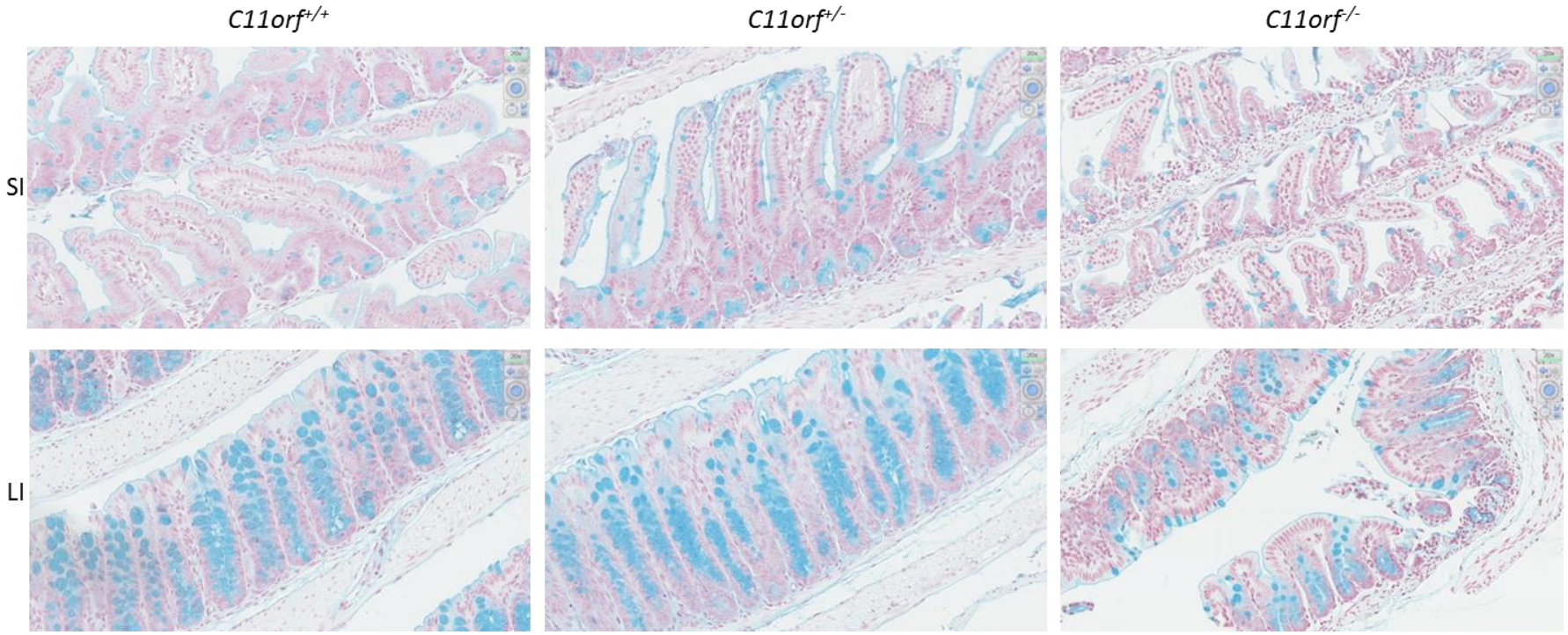


Figure 5.17: FFPE sections of intestinal tissue stained with alcian blue to detect acid-simple non-sulphated and acid-complex mesenchymal mucins. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

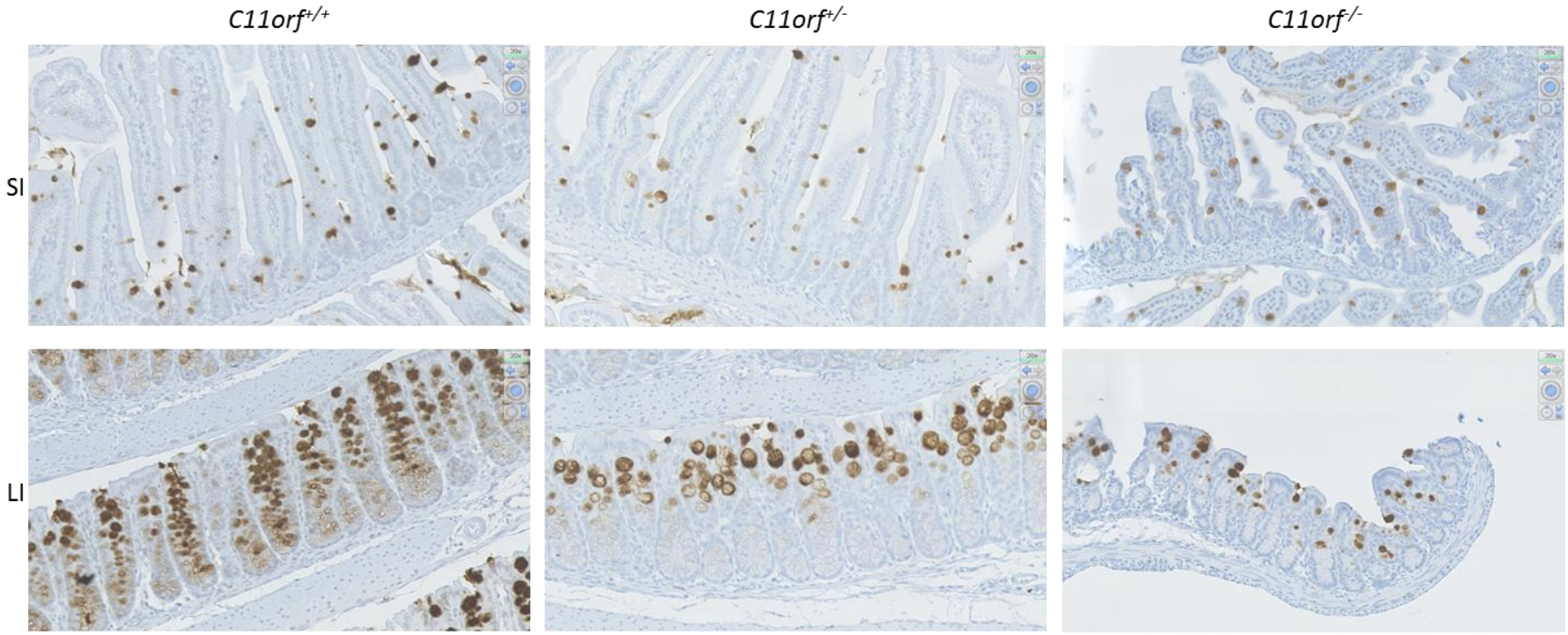


Figure 5.18: FFPE sections of intestinal tissue stained for goblet cell marker Clca1. Tissue is from three-week old *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice. Tissue was washed in PBS and fixed for 24 hours in 10% NBF at room temperature before being processed and paraffin embedded. Images taken at 200X.

5.6 Genome-wide mRNA expression within intestinal tissue

5.6.1 Methods

I harvested and cleaned intestinal tissue from freshly-culled seven-week old female mice. I split the intestine into proximal small intestine (PSI), distal small intestine (DSI), proximal large intestine (PLI), distal large intestine (DLI) and the rectum/anus (R/A). I extracted mRNA using phenol-choroform extraction. This mRNA was sequenced by polyA selection using a NextSeq machine at 25 million paired-end reads by the Wellcome Trust Clinical Research Facility. The sequencing data was subsequently aligned and normalised by Alison Meynert, and differential expression analysis was performed by Graeme Grimes, both from the IGMM Bioinformatics Analysis Core. In addition to the lists of significantly altered genes, I used the normalised counts to look at the expression patterns of individual genes of interest.

Due to the low numbers of *C11orf*$^{-/-}$ mice available for study, the following results are based on only a single mouse per genotype. For preliminary analysis purposes the PSI and DSI results were pooled into SI, and the PLI, DLI and R/A results were pooled into LI, although they are not true replicates. Further work is being carried out by Vidya Rajasekaran and Marion Bacou to gather data from more animals, including males.

5.6.2 Results

5.6.2.1 Comparisons between genotypes and tissues

Differential expression analysis calculated the fold change in expression between each genotype pair – *C11orf*$^{-/-}$ vs *C11orf*$^{+/+}$, *C11orf*$^{-/-}$ vs *C11orf*$^{+/-}$ and *C11orf*$^{+/-}$ vs *C11orf*$^{+/+}$ – within the SI and the LI. The genes were ranked by adjusted p-value (table 5.5-5.10).

I then compared the lists of significantly altered genes in the SI and LI for each genotype comparison (figure 5.19). In all genotype pairs, there are more genes unique to each tissue than shared between SI and LI. This is not surprising, as there are many differences between SI and LI, such as a thicker mucus layer in the LI and higher density of immune cells in the SI (Bowcutt et al., 2014). Additionally, the 11q23.1 locus is only associated with LI cancers and not SI cancers, so we would expect to see some differences between the two in the knockout model. All genes that are significantly altered in both SI and LI are either upregulated in both

or downregulated in both, so while the magnitude of the genes' activity is variable between SI and LI, they do not have opposing roles in the two different tissues.

I also compared the lists of significantly altered genes in each genotype comparison for the SI and the LI (figure 5.20). In both tissues the *C11orf*$^{-/-}$ vs *C11orf*$^{+/+}$ comparison contained more significant genes than either of the other two genotype pairs. From these venn diagrams, and using the direction of fold change for each gene, there are 40 genes in the SI and 14 in the LI where *C11orf*$^{+/-}$ is intermediate to the other two genotypes, meaning expression varies with the dosage of the *C11orf* genes. Nine genes in the SI and 11 in the LI show similar expression for *C11orf*$^{-/-}$ and *C11orf*$^{+/-}$, so both copies of the *C11orf* genes are required to maintain expression levels of these genes. Seven genes in the SI and nine in the LI have similar expression for *C11orf*$^{+/-}$ and *C11orf*$^{+/+}$, so only one copy of the *C11orf* genes needs to be present to maintain expression levels.

There are 11 genes in the SI and nine in the LI where *C11orf*$^{+/-}$ has a more pronounced fold change than *C11orf*$^{-/-}$. Eight genes in the SI and seven in the LI have similar expression for *C11orf*$^{-/-}$ and *C11orf*$^{+/+}$ but not for *C11orf*$^{+/-}$, and there are 16 genes in the SI and 15 in the LI where the fold changes for *C11orf*$^{-/-}$ and *C11orf*$^{+/-}$ compared to *C11orf*$^{+/+}$ are in opposite directions. These genes may be involved in the heterozygous advantage observed in the genotype frequencies. Examples of some of these expression patterns are shown in figure 5.21.

Table 5.5: (a) Upregulated and (b) downregulated genes in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.5a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000096580	*Igkv1-132*	4.000504	1.16E-09	3.88E-06
ENSMUSG00000038393	*Txnip*	0.771696	4.24E-07	0.000504
ENSMUSG00000096206	*Gm22317*	4.295699	4.26E-07	0.000504
ENSMUSG00000093815	*Gm26444*	4.295699	4.26E-07	0.000504
ENSMUSG00000096214	*Gm22634*	4.296132	4.24E-07	0.000504
ENSMUSG00000095969	*Rnu1a1*	4.296132	4.24E-07	0.000504
ENSMUSG00000094826	*Gm23804*	4.296132	4.24E-07	0.000504
ENSMUSG00000096659	*Gm25679*	4.296132	4.24E-07	0.000504
ENSMUSG00000096205	*Gm22068*	4.296132	4.24E-07	0.000504
ENSMUSG00000064856	*Gm23444*	4.105354	6.38E-07	0.000558
ENSMUSG00000094050	*Gm23472*	4.105354	6.38E-07	0.000558
ENSMUSG00000099250	*Gm26191*	4.237009	5.97E-07	0.000558
ENSMUSG00000099021	*Gm22405*	4.237033	5.97E-07	0.000558
ENSMUSG00000076258	*Gm23935*	4.194404	7.30E-07	0.000565
ENSMUSG00000076281	*Gm24270*	4.194456	7.30E-07	0.000565
ENSMUSG00000087943	*Gm24245*	4.194627	7.29E-07	0.000565
ENSMUSG00000088088	*Rmrp*	4.175727	8.86E-07	0.00066
ENSMUSG00000065767	*Gm23849*	3.9081	4.34E-06	0.002647
ENSMUSG00000064702	*Gm24950*	3.9081	4.34E-06	0.002647
ENSMUSG00000065944	*Rnu2-10*	3.9081	4.34E-06	0.002647
5.5b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000044694	*2010007H06Rik*	-5.10397	4.71E-21	9.47E-17
ENSMUSG00000076665	*Ighv7-1*	-4.97152	2.50E-13	2.52E-09
ENSMUSG00000076646	*Ighv2-6-8*	-5.80358	6.84E-13	4.59E-09
ENSMUSG00000106025	*Gm42940*	-2.36275	6.48E-11	3.26E-07
ENSMUSG00000096452	*Ighv1-77*	-3.98198	3.26E-10	1.31E-06
ENSMUSG00000076671	*Ighv13-1*	-3.87739	1.29E-07	0.000325
ENSMUSG00000093955	*Ighv1-34*	-3.81985	1.28E-07	0.000325
ENSMUSG00000054641	*Mmrn1*	-3.60255	1.80E-07	0.000401
ENSMUSG00000052631	*Sh2d6*	-3.89891	5.79E-07	0.000558
ENSMUSG00000076523	*Igkv15-103*	-3.03602	5.65E-07	0.000558
ENSMUSG00000094491	*Igkv1-133*	-3.88721	1.30E-06	0.000937
ENSMUSG00000021867	*Tmem254b*	-1.17836	3.58E-06	0.002485
ENSMUSG00000094951	*Ighv5-6*	-2.8445	2.01E-05	0.009886
ENSMUSG00000094940	*Ighv1-84*	-2.49211	6.99E-05	0.028689
ENSMUSG00000036027	*1810046K07Rik*	-3.16194	7.62E-05	0.029506
ENSMUSG00000096006	*Gm21596*	-2.52091	9.82E-05	0.036589

Table 5.6: (a) Upregulated and (b) downregulated genes in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.6a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000076688	*Ighv15-2*	4.804936	1.05E-10	4.84E-07
ENSMUSG00000096638	*Ighv2-9*	4.299691	1.13E-09	4.46E-06
ENSMUSG00000094322	*Ighv9-4*	3.288061	8.00E-08	0.00017
ENSMUSG00000094319	*Igkv4-54*	4.141772	1.81E-07	0.000345
ENSMUSG00000089911	*Hiat1*	0.788464	1.32E-06	0.001915
ENSMUSG00000038393	*Txnip*	0.545223	1.20E-05	0.013232
ENSMUSG00000043911	*Olfr922*	3.706281	1.29E-05	0.013232
ENSMUSG00000098021	*Gm9522*	2.695	1.46E-05	0.014386
ENSMUSG00000095571	*Ighv5-17*	2.395451	2.11E-05	0.020101
ENSMUSG00000047632	*Fgfbp3*	0.872673	2.49E-05	0.022176
ENSMUSG00000078377	*Gm4294*	3.310395	3.60E-05	0.029175
5.6b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000054641	*Mmrn1*	-7.03663	7.63E-35	2.10E-30
ENSMUSG00000106025	*Gm42940*	-2.97723	1.74E-29	2.40E-25
ENSMUSG00000044694	*2010007H06Rik*	-5.11863	1.13E-27	1.04E-23
ENSMUSG00000076523	*Igkv15-103*	-3.73619	1.02E-12	7.00E-09
ENSMUSG00000096452	*Ighv1-77*	-4.35894	4.75E-11	2.62E-07
ENSMUSG00000034330	*Plcg2*	-0.96762	1.94E-09	6.70E-06
ENSMUSG00000073125	*Xlr3b*	-2.96502	6.05E-09	1.85E-05
ENSMUSG00000095442	*Ighv1-4*	-3.04911	6.90E-09	1.90E-05
ENSMUSG00000051079	*Rgs13*	-4.17152	1.25E-08	3.14E-05
ENSMUSG00000094561	*Ighv1-22*	-3.06901	6.22E-08	0.000143
ENSMUSG00000009246	*Trpm5*	-3.52297	1.88E-07	0.000345
ENSMUSG00000021867	*Tmem254b*	-1.0399	3.19E-07	0.000549
ENSMUSG00000020901	*Pik3r5*	-1.70761	5.79E-07	0.00094
ENSMUSG00000094940	*Ighv1-84*	-2.76271	8.96E-07	0.001373
ENSMUSG00000094345	*Igkv14-126*	-3.35858	1.64E-06	0.002259
ENSMUSG00000076598	*Igkv3-7*	-2.51215	2.31E-06	0.00303
ENSMUSG00000076671	*Ighv13-1*	-3.00713	7.40E-06	0.00928
ENSMUSG00000025889	*Snca*	-3.1989	9.56E-06	0.011459
ENSMUSG00000019429	*Ffar3*	-2.45661	1.29E-05	0.013232
ENSMUSG00000033439	*Trmt13*	-0.96867	1.30E-05	0.013232

Table 5.7: (a) Upregulated and (b) downregulated genes in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/-}$ SI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.7a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUST00000214998	*March5-205*	4.145081	6.62E-10	4.12E-06
ENSMUSG00000086377	*4930529C04Rik*	4.169928	3.85E-07	0.001062
ENSMUSG00000064856	*Gm23444*	4.104652	6.51E-07	0.001471
ENSMUSG00000094050	*Gm23472*	4.104652	6.51E-07	0.001471
ENSMUSG00000030223	*Ptpro*	1.499331	4.16E-06	0.006904
ENSMUSG00000076652	*Ighv7-3*	2.885547	7.27E-06	0.009037
ENSMUSG00000076525	*Igkv1-99*	3.561475	7.81E-06	0.009245
ENSMUSG00000020659	*Cbll1*	0.759941	1.84E-05	0.01908
ENSMUSG00000084128	*Esrp2*	0.416835	3.42E-05	0.03036
ENSMUSG00000079343	*C1s2*	3.520282	3.41E-05	0.03036
ENSMUSG00000101166	*Gm28496*	3.423767	3.84E-05	0.032915
ENSMUSG00000103889	*RP23-347F2.2*	2.947448	4.18E-05	0.03468
ENSMUSG00000006191	*Cdkal1*	0.894438	4.44E-05	0.035587
ENSMUSG00000104768	*Gm38412*	2.533581	5.10E-05	0.039593
5.7b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000076547	*Igkv4-70*	-4.29404	8.14E-15	2.02E-10
ENSMUSG00000044694	*2010007H06Rik*	-3.89452	9.89E-13	1.23E-08
ENSMUSG00000095633	*Igkv4-58*	-4.28741	6.30E-12	5.22E-08
ENSMUSG00000054641	*Mmrn1*	-4.1298	2.19E-09	1.09E-05
ENSMUSG00000094087	*Ighv1-61*	-4.21319	3.79E-09	1.57E-05
ENSMUSG00000035299	*Mid1*	-1.76366	5.92E-09	2.10E-05
ENSMUSG00000095338	*Igkv3-9*	-4.19323	5.55E-08	0.000172
ENSMUSG00000096577	*Ighv1-71*	-3.97242	7.48E-07	0.001549
ENSMUSG00000076534	*Igkv12-89*	-3.42182	3.40E-06	0.006498
ENSMUSG00000017734	*Dbndd2*	-0.69298	3.74E-06	0.006635
ENSMUSG00000076598	*Igkv3-7*	-2.15777	4.93E-06	0.007659
ENSMUSG00000034871	*Fam151a*	-3.74887	5.47E-06	0.007989
ENSMUSG00000096670	*Ighv2-6*	-3.02718	5.78E-06	0.007989
ENSMUSG00000076594	*Igkv6-13*	-2.80214	6.98E-06	0.009037
ENSMUSG00000096844	*Igkv6-14*	-2.45214	1.13E-05	0.012772
ENSMUSG00000095285	*Ighv5-9*	-2.8867	1.52E-05	0.01641
ENSMUSG00000094262	*Igkv4-62*	-3.53369	2.30E-05	0.022838
ENSMUST00000202978	*Men1-224*	-2.8061	2.53E-05	0.024234
ENSMUSG00000025889	*Snca*	-3.06048	5.48E-05	0.041277

Table 5.8: (a) Upregulated and (b) downregulated genes in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/-}$ LI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.8a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000076665	*Ighv7-1*	4.468703	1.29E-13	6.74E-10
ENSMUSG00000094322	*Ighv9-4*	4.004452	6.88E-11	2.26E-07
ENSMUSG00000094319	*Igkv4-54*	4.843086	1.39E-09	3.09E-06
ENSMUSG00000096638	*Ighv2-9*	3.106664	7.13E-06	0.009343
ENSMUSG00000047632	*Fgfbp3*	0.878853	9.83E-06	0.011721
ENSMUSG00000094502	*Ighv1-69*	2.806439	1.77E-05	0.020179
ENSMUSG00000038393	*Txnip*	0.503153	5.32E-05	0.048925
ENSMUSG00000064352	*mt-Ts1*	2.073457	5.50E-05	0.048925
ENSMUSG00000048489	*8430408G22Rik*	2.121217	5.72E-05	0.048925
5.8b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000054641	*Mmrn1*	-7.91474	1.00E-43	2.63E-39
ENSMUSG00000103254	*Ighv1-15*	-4.31451	7.89E-16	6.90E-12
ENSMUSG00000035299	*Mid1*	-1.99057	7.90E-16	6.90E-12
ENSMUSG00000044694	*2010007H06Rik*	-3.75426	1.97E-15	1.29E-11
ENSMUSG00000096461	*Igkv14-130*	-4.86508	1.02E-11	3.82E-08
ENSMUSG00000076547	*Igkv4-70*	-3.23436	1.01E-11	3.82E-08
ENSMUSG00000094087	*Ighv1-61*	-4.19307	1.09E-10	3.19E-07
ENSMUSG00000095633	*Igkv4-58*	-3.33389	1.41E-09	3.09E-06
ENSMUSG00000076598	*Igkv3-7*	-3.15956	1.29E-09	3.09E-06
ENSMUSG00000095335	*Igkv3-5*	-3.78311	6.59E-09	1.33E-05
ENSMUSG00000025889	*Snca*	-3.86558	4.83E-08	9.05E-05
ENSMUSG00000093894	*Ighv1-53*	-2.75943	1.85E-07	0.000324
ENSMUSG00000095338	*Igkv3-9*	-3.78214	2.68E-07	0.000439
ENSMUSG00000076538	*Igkv13-84*	-3.65861	4.93E-07	0.00076
ENSMUSG00000096715	*Igkv3-4*	-3.48415	9.69E-07	0.001412
ENSMUSG00000041859	*Mcm3*	-0.81427	6.49E-06	0.008952
ENSMUSG00000096078	*Ighv1-62-2*	-3.26719	8.47E-06	0.010578
ENSMUSG00000087263	*Gm15726*	-1.82762	2.49E-05	0.027221
ENSMUSG00000095285	*Ighv5-9*	-2.44515	3.47E-05	0.036394
ENSMUSG00000035049	*Rrp12*	-0.66645	3.68E-05	0.037157

Table 5.9: (a) Upregulated and (b) downregulated genes in *C11orf*$^{+/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.9a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000096670	*Ighv2-6*	5.834975	2.06E-17	2.50E-13
ENSMUSG00000076594	*Igkv6-13*	5.259002	7.94E-17	6.40E-13
ENSMUSG00000076547	*Igkv4-70*	4.451957	3.67E-16	2.22E-12
ENSMUSG00000035299	*Mid1*	1.702836	1.47E-08	3.55E-05
ENSMUSG00000094087	*Ighv1-61*	3.789115	7.52E-08	0.000151
ENSMUSG00000076534	*Igkv12-89*	3.946155	8.13E-08	0.000151
ENSMUSG00000095335	*Igkv3-5*	3.780899	2.58E-07	0.000446
ENSMUSG00000076670	*Ighv3-5*	3.942184	4.77E-07	0.000769
ENSMUSG00000095633	*Igkv4-58*	3.059841	5.20E-07	0.000787
ENSMUSG00000094262	*Igkv4-62*	4.10988	1.12E-06	0.00159
ENSMUSG00000076548	*Igkv4-69*	3.901295	2.23E-06	0.002993
ENSMUSG00000102524	*Ighv1-2*	3.586638	1.18E-05	0.013019
5.9b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000076665	*Ighv7-1*	-6.06438	4.53E-19	1.10E-14
ENSMUSG00000076646	*Ighv2-6-8*	-6.06414	3.62E-14	1.75E-10
ENSMUSG00000106025	*Gm42940*	-2.25281	1.89E-11	7.63E-08
ENSMUST00000214998	*March5-205*	-4.25946	1.48E-10	5.10E-07
ENSMUSG00000094491	*Igkv1-133*	-5.15148	1.70E-10	5.15E-07
ENSMUSG00000093955	*Ighv1-34*	-4.28828	2.80E-09	7.53E-06
ENSMUSG00000095771	*Igkv14-111*	-3.0836	4.04E-08	8.88E-05
ENSMUSG00000096452	*Ighv1-77*	-2.89367	2.86E-06	0.003495
ENSMUSG00000076523	*Igkv15-103*	-2.83776	2.89E-06	0.003495
ENSMUSG00000076671	*Ighv13-1*	-3.34672	4.26E-06	0.004905
ENSMUSG00000076571	*Igkv5-37*	-3.28667	1.89E-05	0.019904
ENSMUSG00000025316	*Banp*	-0.75399	2.09E-05	0.021025
ENSMUSG00000076680	*Ighv6-6*	-3.19045	3.22E-05	0.031114
ENSMUSG00000094930	*Igkv6-25*	-2.64835	3.73E-05	0.034737
ENSMUSG00000094951	*Ighv5-6*	-2.70976	4.54E-05	0.040705
ENSMUSG00000076525	*Igkv1-99*	-3.24042	4.72E-05	0.040772

Table 5.10: (a) Upregulated and (b) downregulated genes in *C11orf*$^{+/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with adjusted p-value lower than 0.05 are included; only the top 20 results are shown if more than that were significant, with full results in appendix 2.

5.10a) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000035299	*Mid1*	2.168241	2.28E-18	1.58E-14
ENSMUSG00000076547	*Igkv4-70*	4.49414	6.87E-18	3.57E-14
ENSMUSG00000095335	*Igkv3-5*	5.168307	6.82E-15	2.83E-11
ENSMUSG00000095633	*Igkv4-58*	4.003953	6.74E-12	2.34E-08
ENSMUSG00000103254	*Ighv1-15*	3.181438	2.50E-09	5.78E-06
ENSMUSG00000094051	*Ighv1-36*	3.544269	6.43E-08	0.000112
ENSMUSG00000094087	*Ighv1-61*	3.49377	7.64E-08	0.000122
ENSMUSG00000089911	*Hiat1*	0.863741	1.17E-07	0.000165
ENSMUSG00000078377	*Gm4294*	4.237886	1.19E-07	0.000165
ENSMUSG00000096577	*Ighv1-71*	3.355272	6.93E-06	0.006267
ENSMUSG00000094117	*Igkv3-12*	2.58294	1.46E-05	0.012626
ENSMUSG00000037386	*Rims2*	1.568707	3.72E-05	0.029767
5.10b) Ensembl gene ID	**Gene name**	**Log_2 fold change**	**p-value**	**Adjusted p-value**
ENSMUSG00000106025	*Gm42940*	-3.06997	2.18E-30	4.53E-26
ENSMUSG00000076665	*Ighv7-1*	-5.77113	1.08E-21	1.12E-17
ENSMUSG00000021867	*Tmem254b*	-1.31436	1.32E-10	3.92E-07
ENSMUSG00000094345	*Igkv14-126*	-4.30699	1.42E-09	3.69E-06
ENSMUSG00000095442	*Ighv1-4*	-3.00972	1.19E-08	2.47E-05
ENSMUSG00000076523	*Igkv15-103*	-2.82879	6.33E-08	0.000112
ENSMUSG00000076680	*Ighv6-6*	-3.49247	4.66E-07	0.000605
ENSMUSG00000096921	*Gm26822*	-3.70606	7.67E-07	0.000938
ENSMUSG00000094561	*Ighv1-22*	-2.68888	2.07E-06	0.002388
ENSMUSG00000096452	*Ighv1-77*	-2.7362	3.22E-06	0.003528
ENSMUSG00000072676	*Tmem254a*	-1.32681	4.03E-06	0.004187
ENSMUSG00000094491	*Igkv1-133*	-3.45514	4.40E-06	0.004359
ENSMUSG00000076646	*Ighv2-6-8*	-3.41646	6.57E-06	0.006211
ENSMUSG00000033439	*Trmt13*	-0.9466	2.03E-05	0.016862
ENSMUSG00000076940	*Iglv2*	-2.8671	3.91E-05	0.030094

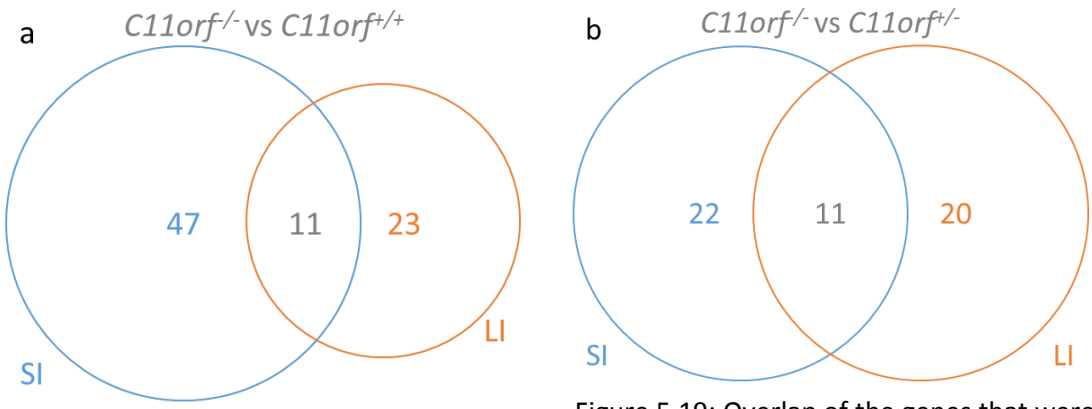


Figure 5.19: Overlap of the genes that were significantly altered ($p < 0.05$) in mRNA-seq analysis between genotypes, showing how many genes were significantly altered in SI, LI or both. Venn diagrams show the overlap between tissues when comparing (a) homozygous knockout and wild type, (b) homozygous knockout and heterozygous knockout, and (c) heterozygous knockout and wild type.

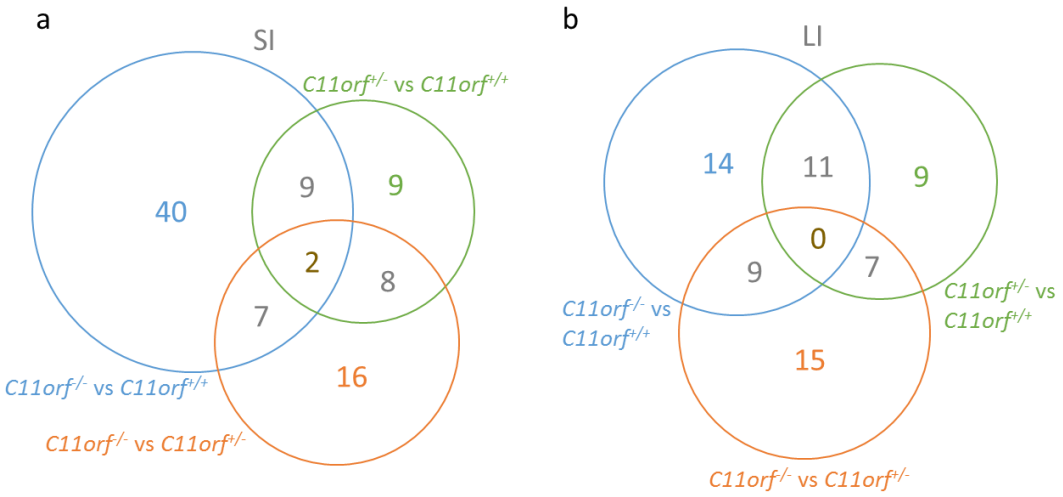


Figure 5.20: Overlap of the genes that were significantly altered ($p < 0.05$) in mRNA-seq analysis between tissues, showing how many genes were significantly altered in one, two or three genotype comparisons. Venn diagrams show the overlap in genes in (a) the SI and (b) the LI.

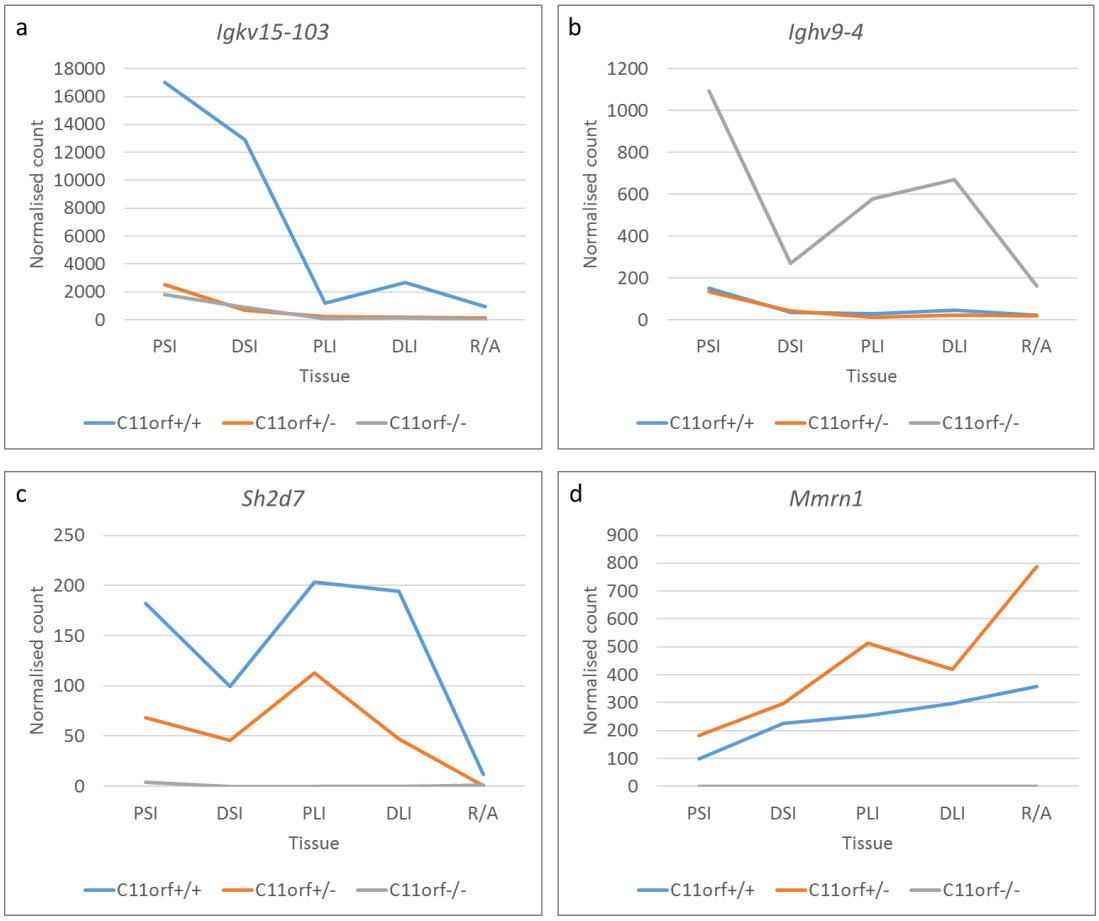


Figure 5.21: Different expression patterns seen in the mRNA-seq data, with the normalised counts in each genotype (*C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$) shown for each tissue analysed (proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and rectum/anus). All of these genes were significantly differentially expressed at the adjusted p-value < 0.05 threshold. (a) Example of a gene that is downregulated in both heterozygous and homozygous knockouts. (b) Example of a gene that is upregulated in homozygous knockouts but not heterozygous knockouts. (c) Example of a gene showing partial downregulation in heterozygous knockouts and total downregulation in homozygous knockouts. (d) Example of a gene that is upregulated in heterozygous knockouts and downregulated in homozygous knockouts.

5.6.2.2 Gene ontology enrichment analysis

I used the online GOrilla tool (www.cbl-gorilla.cs.technion.ac.il, (Eden et al., 2007, 2009)) to run gene ontology (GO) enrichment analysis on the resulting lists of upregulated and downregulated genes for each genotype comparison in SI and LI, with the gene lists being ranked by adjusted p-value (tables 5.11-5.22).

Many of the enriched GO terms are immune-related, such as immunoglobin complex, complement activation, phagocytosis and B-cell receptor signalling. There is no clear pattern of what is upregulated and what is downregulated, with immune-related terms being common in both, and some individual terms being listed as both upregulated and downregulated in the same comparison. Many of the significantly altered genes, both upregulated and downregulated, were immunoglobin chains, and this is presumably the major factor in the strong immune association, although there are also some other immune-related genes. These include *Rgs13*, a regulator of G protein signalling with roles in mast cell and germinal centre B lymphocyte responses (Bansal et al., 2008; Shi et al., 2002), and *Sh2d7*, a paralogue of *Sh2d2a*, which is a T cell regulator linked to arthritis (Smerdel et al., 2004).

Cellular component terms are largely membrane-bound, including the nucleus, endoplasmic reticulum and Golgi apparatus alongside more generic membrane component terms.

Table 5.11: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.11a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0002377	immunoglobulin production	5.13E-17	6.53E-13	35.88 (10130,67,59,14)
GO:0002440	production of molecular mediator of immune response	5.97E-16	3.80E-12	30.82 (10130,78,59,14)
GO:0006955	immune response	3.02E-13	1.28E-09	7.29 (10130,543,64,25)
GO:0043170	macromolecule metabolic process	5.09E-13	1.62E-09	1.32 (10130,2984,1232,480)
GO:0006807	nitrogen compound metabolic process	3.64E-12	9.27E-09	1.28 (10130,3353,1232,524)
GO:0044237	cellular metabolic process	1.80E-11	3.81E-08	1.26 (10130,3538,1234,545)
GO:0006910	phagocytosis, recognition	4.97E-11	9.04E-08	21.23 (10130,82,64,11)
GO:0006958	complement activation, classical pathway	8.60E-11	1.22E-07	20.25 (10130,86,64,11)
GO:0071704	organic substance metabolic process	8.30E-11	1.32E-07	1.24 (10130,3776,1232,571)
GO:0006911	phagocytosis, engulfment	1.63E-10	2.08E-07	19.13 (10130,91,64,11)
GO:0006956	complement activation	1.81E-10	2.10E-07	18.92 (10130,92,64,11)
GO:0050853	B cell receptor signaling pathway	2.03E-10	2.16E-07	18.72 (10130,93,64,11)
GO:0099024	plasma membrane invagination	2.62E-10	2.57E-07	18.33 (10130,95,64,11)
GO:0010324	membrane invagination	3.31E-10	3.01E-07	17.95 (10130,97,64,11)
GO:0072376	protein activation cascade	4.72E-10	3.53E-07	17.41 (10130,100,64,11)
5.11b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0005488	binding	3.49E-11	6.13E-08	1.16 (10130,6038,1099,762)
GO:0034987	immunoglobulin receptor binding	2.08E-11	7.32E-08	22.91 (10130,76,64,11)
GO:0003823	antigen binding	9.93E-10	1.16E-06	16.27 (10130,107,64,11)

GO:0016740	transferase activity	1.07E-07	9.37E-05	1.47 (10130,1047,1234,188)
GO:0003676	nucleic acid binding	4.68E-07	3.29E-04	1.38 (10130,1461,1168,233)
GO:1901363	heterocyclic compound binding	6.69E-07	3.92E-04	1.27 (10130,2343,1234,363)
GO:0097159	organic cyclic compound binding	1.36E-06	6.83E-04	1.26 (10130,2380,1234,366)
GO:0005515	protein binding	1.83E-06	8.02E-04	1.18 (10130,4120,1142,548)
GO:0019899	enzyme binding	9.34E-06	3.65E-03	1.39 (10130,1061,1253,183)
GO:0043167	ion binding	8.02E-05	2.82E-02	1.21 (10130,2513,1248,375)
5.11c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0044424	intracellular part	4.30E-21	7.11E-18	1.21 (10130,6247,1244,928)
GO:0043229	intracellular organelle	8.04E-18	4.43E-15	1.25 (10130,5135,1229,778)
GO:0043226	organelle	6.93E-18	5.73E-15	1.24 (10130,5234,1229,790)
GO:0043227	membrane-bounded organelle	4.10E-17	1.69E-14	1.26 (10130,4746,1255,742)
GO:0043231	intracellular membrane-bounded organelle	6.76E-17	2.23E-14	1.28 (10130,4435,1228,689)
GO:0044428	nuclear part	4.33E-16	1.19E-13	1.52 (10130,1893,1225,349)
GO:0044446	intracellular organelle part	6.43E-15	1.52E-12	1.33 (10130,3356,1254,554)
GO:0044422	organelle part	7.25E-14	1.50E-11	1.31 (10130,3481,1254,566)
GO:0005634	nucleus	9.43E-14	1.73E-11	1.35 (10130,2730,1233,450)
GO:0044464	cell part	5.73E-13	9.46E-11	1.13 (10130,7318,1205,987)
GO:0005654	nucleoplasm	4.85E-12	7.29E-10	1.63 (10130,965,1254,195)
GO:0042571	immunoglobulin complex, circulating	1.10E-11	1.51E-09	24.18 (10130,72,64,11)
GO:0019814	immunoglobulin complex	1.74E-11	2.22E-09	23.21 (10130,75,64,11)
GO:1902494	catalytic complex	2.21E-08	2.61E-06	1.64 (10130,660,1253,134)
GO:1990234	transferase complex	4.56E-08	4.71E-06	1.88 (10130,356,1253,83)

Table 5.12: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.12a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0002377	immunoglobulin production	2.38E-16	3.09E-12	12.65 (9159,67,227,21)
GO:0002440	production of molecular mediator of immune response	3.63E-15	2.36E-11	11.30 (9159,75,227,21)
GO:0006955	immune response	7.76E-14	3.36E-10	4.20 (9159,414,232,44)
GO:0050853	B cell receptor signaling pathway	2.94E-13	9.54E-10	21.20 (9159,54,104,13)
GO:0006910	phagocytosis, recognition	5.25E-13	9.74E-10	14.43 (9159,46,207,15)
GO:0006956	complement activation	4.87E-13	1.05E-09	8.01 (9159,60,400,21)
GO:0006958	complement activation, classical pathway	4.32E-13	1.12E-09	13.11 (9159,54,207,16)
GO:0006911	phagocytosis, engulfment	1.52E-12	2.47E-09	12.21 (9159,58,207,16)
GO:0072376	protein activation cascade	3.12E-12	4.50E-09	7.40 (9159,65,400,21)
GO:0099024	plasma membrane invagination	6.20E-12	8.05E-09	11.24 (9159,63,207,16)
GO:0002376	immune system process	1.26E-11	1.49E-08	3.93 (9159,686,119,35)
GO:0010324	membrane invagination	1.85E-11	2.00E-08	10.57 (9159,67,207,16)
GO:0050851	antigen receptor-mediated signaling pathway	1.43E-10	1.43E-07	13.63 (9159,84,104,13)
GO:0050871	positive regulation of B cell activation	1.60E-10	1.48E-07	58.55 (9159,73,15,7)
GO:0009617	response to bacterium	2.23E-10	1.93E-07	8.76 (9159,190,88,16)
5.12b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	5.25E-13	1.88E-09	14.43 (9159,46,207,15)
GO:0003823	antigen binding	1.92E-11	3.43E-08	7.31 (9159,69,363,20)
GO:0015267	channel activity	1.30E-07	9.32E-05	2.29 (9159,202,970,49)

GO:0022838	substrate-specific channel activity	1.94E-07	1.15E-04	2.31 (9159,192,970,47)
GO:0022803	passive transmembrane transporter activity	1.30E-07	1.16E-04	2.29 (9159,202,970,49)
GO:0005216	ion channel activity	1.01E-07	1.20E-04	2.35 (9159,189,970,47)
GO:0005261	cation channel activity	3.83E-07	1.96E-04	2.58 (9159,139,970,38)
GO:0022836	gated channel activity	1.34E-06	5.99E-04	2.47 (9159,145,970,38)
GO:0022839	ion gated channel activity	5.31E-06	2.11E-03	2.41 (9159,141,970,36)
GO:0022890	inorganic cation transmembrane transporter activity	2.33E-05	8.32E-03	2.23 (9159,264,575,37)
GO:0005516	calmodulin binding	3.83E-05	1.05E-02	2.54 (9159,94,1037,27)
GO:0004930	G-protein coupled receptor activity	3.61E-05	1.07E-02	2.16 (9159,161,947,36)
GO:0015318	inorganic molecular entity transmembrane transporter activity	3.35E-05	1.09E-02	1.72 (9159,370,992,69)
GO:0015075	ion transmembrane transporter activity	4.86E-05	1.24E-02	1.87 (9159,404,656,54)
GO:0008324	cation transmembrane transporter activity	7.46E-05	1.78E-02	2.10 (9159,288,575,38)
5.12c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	2.56E-13	2.09E-10	15.08 (9159,44,207,15)
GO:0042571	immunoglobulin complex, circulating	1.74E-13	2.85E-10	15.43 (9159,43,207,15)
GO:0097458	neuron part	2.88E-10	1.57E-07	1.62 (9159,944,1031,172)
GO:0009897	external side of plasma membrane	7.87E-09	3.21E-06	3.79 (9159,192,352,28)
GO:0005886	plasma membrane	4.21E-08	1.37E-05	1.35 (9159,1922,1031,292)
GO:0043005	neuron projection	7.06E-08	1.92E-05	1.65 (9159,668,1031,124)
GO:0098552	side of membrane	1.01E-07	2.36E-05	13.33 (9159,229,27,9)
GO:0044459	plasma membrane part	1.40E-07	2.85E-05	1.45 (9159,1280,984,199)
GO:0120025	plasma membrane bounded cell projection	4.94E-07	8.96E-05	1.51 (9159,930,1003,154)
GO:0031224	intrinsic component of membrane	5.81E-07	9.48E-05	1.29 (9159,2279,1002,322)

GO:0044425	membrane part	7.75E-07	1.15E-04	1.24 (9159,2963,1006,403)
GO:0045202	synapse	1.06E-06	1.44E-04	1.86 (9159,353,1003,72)
GO:0044456	synapse part	2.31E-06	2.90E-04	1.75 (9159,429,1003,82)
GO:0034703	cation channel complex	2.77E-06	3.23E-04	2.66 (9159,106,1006,31)
GO:1902495	transmembrane transporter complex	5.09E-06	5.54E-04	2.29 (9159,147,1006,37)

Table 5.13: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.13a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0050853	B cell receptor signaling pathway	4.30E-10	1.36E-06	54.16 (10073,42,31,7)
GO:0006911	phagocytosis, engulfment	6.25E-10	1.58E-06	51.69 (10073,44,31,7)
GO:0006958	complement activation, classical pathway	4.30E-10	1.81E-06	54.16 (10073,42,31,7)
GO:0006955	immune response	3.26E-10	2.06E-06	11.34 (10073,361,32,13)
GO:0006956	complement activation	1.19E-09	2.50E-06	47.39 (10073,48,31,7)
GO:0099024	plasma membrane invagination	1.87E-09	3.38E-06	44.60 (10073,51,31,7)
GO:0072376	protein activation cascade	2.46E-09	3.46E-06	42.92 (10073,53,31,7)
GO:0050871	positive regulation of B cell activation	2.22E-09	3.50E-06	30.52 (10073,66,40,8)
GO:0006910	phagocytosis, recognition	2.99E-10	3.78E-06	56.86 (10073,40,31,7)
GO:0010324	membrane invagination	3.39E-09	3.89E-06	41.36 (10073,55,31,7)
GO:0050851	antigen receptor-mediated signaling pathway	3.20E-09	4.05E-06	29.20 (10073,69,40,8)
GO:0002429	immune response-activating cell surface receptor signaling pathway	5.07E-09	5.34E-06	27.60 (10073,73,40,8)
GO:0050864	regulation of B cell activation	6.61E-09	6.43E-06	26.86 (10073,75,40,8)

GO:0002768	immune response-regulating cell surface receptor signaling pathway	9.46E-09	8.54E-06	25.83 (10073,78,40,8)
GO:0002757	immune response-activating signal transduction	3.52E-08	2.96E-05	21.90 (10073,92,40,8)
5.13b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	1.63E-10	5.94E-07	61.47 (10073,37,31,7)
GO:0005515	protein binding	7.30E-10	1.33E-06	1.22 (10073,4151,1177,591)
GO:0003823	antigen binding	2.87E-09	3.49E-06	42.12 (10073,54,31,7)
GO:0005488	binding	1.74E-08	1.59E-05	1.14 (10073,5975,1177,794)
5.13c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	1.08E-10	8.74E-08	64.99 (10073,35,31,7)
GO:0042571	immunoglobulin complex, circulating	1.08E-10	1.75E-07	64.99 (10073,35,31,7)
GO:0009897	external side of plasma membrane	5.72E-07	3.09E-04	13.80 (10073,146,40,8)
GO:0098552	side of membrane	1.73E-06	7.01E-04	20.32 (10073,175,17,6)
GO:0044464	cell part	1.04E-04	2.83E-02	1.07 (10073,7518,1180,941)
GO:0015030	Cajal body	1.33E-04	3.09E-02	5.85 (10073,17,810,8)
GO:0044421	extracellular region part	9.86E-05	3.20E-02	7.63 (10073,770,12,7)

Table 5.14: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.14a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0006955	immune response	1.44E-37	1.87E-33	2.90 (9699,588,950,167)

GO:0002376	immune system process	6.89E-36	4.47E-32	2.20 (9699,956,1170,254)
GO:0050853	B cell receptor signaling pathway	9.73E-34	4.20E-30	6.07 (9699,103,946,61)
GO:0006910	phagocytosis, recognition	1.62E-28	4.20E-25	7.46 (9699,86,695,46)
GO:0050864	regulation of B cell activation	1.61E-28	5.22E-25	4.64 (9699,160,914,70)
GO:0050871	positive regulation of B cell activation	3.51E-28	7.59E-25	5.01 (9699,129,946,63)
GO:0006958	complement activation, classical pathway	9.97E-28	1.85E-24	5.78 (9699,94,946,53)
GO:0002694	regulation of leukocyte activation	1.39E-27	2.26E-24	3.16 (9699,371,917,111)
GO:0002377	immunoglobulin production	1.62E-27	2.33E-24	16.43 (9699,77,230,30)
GO:0006956	complement activation	5.90E-27	7.65E-24	5.54 (9699,100,946,54)
GO:0051249	regulation of lymphocyte activation	8.34E-27	9.83E-24	3.32 (9699,322,917,101)
GO:0051251	positive regulation of lymphocyte activation	2.78E-26	3.00E-23	3.75 (9699,237,917,84)
GO:0050851	antigen receptor-mediated signaling pathway	4.48E-26	4.15E-23	4.43 (9699,152,965,67)
GO:0002429	immune response-activating cell surface receptor signaling pathway	4.29E-26	4.28E-23	3.94 (9699,166,1111,75)
GO:0002440	production of molecular mediator of immune response	9.14E-26	7.90E-23	14.71 (9699,86,230,30)
5.14b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0003823	antigen binding	6.32E-31	2.16E-27	6.57 (9699,119,695,56)
GO:0034987	immunoglobulin receptor binding	3.10E-28	5.31E-25	7.57 (9699,83,695,45)
GO:0060089	molecular transducer activity	8.94E-09	7.65E-06	1.55 (9699,927,1185,175)
GO:0038023	signaling receptor activity	7.96E-09	9.08E-06	1.55 (9699,925,1185,175)
GO:0004888	transmembrane signaling receptor activity	2.13E-08	1.46E-05	1.57 (9699,834,1170,158)
GO:0005102	signaling receptor binding	4.03E-08	2.30E-05	1.73 (9699,904,695,112)
GO:0030246	carbohydrate binding	1.47E-06	7.17E-04	2.74 (9699,166,661,31)
GO:0022836	gated channel activity	2.00E-06	8.56E-04	2.20 (9699,179,1083,44)

GO:0022839	ion gated channel activity	2.54E-06	9.65E-04	2.20 (9699,175,1083,43)
GO:0023023	MHC protein complex binding	8.29E-06	2.84E-03	7.73 (9699,8,1098,7)
GO:0023026	MHC class II protein complex binding	9.66E-06	3.01E-03	8.83 (9699,6,1098,6)
GO:0004896	cytokine receptor activity	1.13E-05	3.24E-03	2.92 (9699,64,1194,23)
GO:0015267	channel activity	1.02E-04	2.49E-02	1.83 (9699,245,1083,50)
GO:0005216	ion channel activity	1.12E-04	2.55E-02	1.86 (9699,226,1083,47)
GO:0022803	passive transmembrane transporter activity	1.02E-04	2.68E-02	1.83 (9699,245,1083,50)
5.14c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	4.83E-31	7.89E-28	8.00 (9699,82,695,47)
GO:0042571	immunoglobulin complex, circulating	5.73E-30	4.68E-27	6.57 (9699,78,946,50)
GO:0009897	external side of plasma membrane	5.88E-28	3.20E-25	3.00 (9699,317,1194,117)
GO:0098552	side of membrane	5.86E-26	2.39E-23	3.59 (9699,354,695,91)
GO:0044459	plasma membrane part	7.00E-17	2.29E-14	1.61 (9699,1508,1194,299)
GO:0044425	membrane part	1.36E-16	3.70E-14	1.35 (9699,3309,1173,542)
GO:0031224	intrinsic component of membrane	1.68E-12	3.93E-10	1.35 (9699,2619,1163,423)
GO:0016021	integral component of membrane	9.02E-11	1.84E-08	1.33 (9699,2533,1158,402)
GO:0042613	MHC class II protein complex	1.84E-08	3.34E-06	8.83 (9699,9,1098,9)
GO:0044421	extracellular region part	1.33E-07	2.18E-05	1.53 (9699,853,1183,159)
GO:0005886	plasma membrane	8.16E-07	1.03E-04	1.29 (9699,2156,1161,332)
GO:0005887	integral component of plasma membrane	7.98E-07	1.09E-04	1.61 (9699,589,1186,116)
GO:0031226	intrinsic component of plasma membrane	7.38E-07	1.10E-04	1.59 (9699,627,1186,122)
GO:0042611	MHC protein complex	1.75E-06	2.05E-04	5.72 (9699,17,1098,11)
GO:0097651	phosphatidylinositol 3-kinase complex, class I	8.63E-06	8.82E-04	334.45 (9699,2,29,2)

Table 5.15: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/-}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.15a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0044237	cellular metabolic process	8.50E-14	1.06E-09	1.33 (9133,3018,1071,470)
GO:0008152	metabolic process	5.48E-13	3.43E-09	1.28 (9133,3453,1087,527)
GO:0044238	primary metabolic process	5.93E-12	2.47E-08	1.29 (9133,3157,1071,479)
GO:0071704	organic substance metabolic process	1.98E-11	6.20E-08	1.28 (9133,3283,1071,492)
GO:0006807	nitrogen compound metabolic process	8.85E-11	1.84E-07	1.31 (9133,2926,962,405)
GO:0034641	cellular nitrogen compound metabolic process	7.53E-11	1.88E-07	1.41 (9133,1760,1139,309)
GO:0046483	heterocycle metabolic process	4.87E-10	7.62E-07	1.46 (9133,1633,957,249)
GO:1901360	organic cyclic compound metabolic process	4.77E-10	8.52E-07	1.44 (9133,1739,957,262)
GO:0006139	nucleobase-containing compound metabolic process	9.40E-10	1.31E-06	1.41 (9133,1578,1139,278)
GO:0006725	cellular aromatic compound metabolic process	2.43E-09	3.04E-06	1.43 (9133,1658,957,249)
GO:0034660	ncRNA metabolic process	5.48E-09	6.23E-06	2.74 (9133,154,909,42)
GO:0022613	ribonucleoprotein complex biogenesis	1.06E-08	1.10E-05	6.03 (9133,25,909,15)
GO:0090304	nucleic acid metabolic process	1.38E-08	1.33E-05	1.46 (9133,1391,962,214)
GO:0006955	immune response	1.67E-08	1.49E-05	5.92 (9133,495,53,17)
GO:0006396	RNA processing	3.32E-08	2.77E-05	2.18 (9133,270,932,60)
5.15b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0003723	RNA binding	4.00E-07	1.35E-03	1.86 (9133,402,940,77)
GO:0005488	binding	1.45E-06	2.46E-03	1.13 (9133,5405,1057,705)
GO:0034987	immunoglobulin receptor binding	1.04E-05	5.87E-03	18.05 (9133,66,46,6)

GO:0003824	catalytic activity	5.93E-06	6.69E-03	1.25 (9133,2347,1087,349)
GO:1901363	heterocyclic compound binding	1.01E-05	6.86E-03	1.26 (9133,2143,1090,322)
GO:0097159	organic cyclic compound binding	9.62E-06	8.13E-03	1.26 (9133,2181,1090,327)
GO:0003823	antigen binding	5.10E-05	2.47E-02	13.85 (9133,86,46,6)
GO:0003676	nucleic acid binding	6.20E-05	2.62E-02	1.36 (9133,1382,848,175)
5.15c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0044424	intracellular part	2.77E-32	4.25E-29	1.29 (9133,5374,1071,814)
GO:0043231	intracellular membrane-bounded organelle	8.00E-32	6.14E-29	1.44 (9133,3705,1070,623)
GO:0043226	organelle	5.69E-31	2.91E-28	1.36 (9133,4465,1070,710)
GO:0043229	intracellular organelle	5.89E-30	2.26E-27	1.36 (9133,4368,1070,696)
GO:0043227	membrane-bounded organelle	1.11E-28	3.42E-26	1.38 (9133,4023,1070,652)
GO:0044446	intracellular organelle part	5.34E-24	1.37E-21	1.50 (9133,2732,1042,467)
GO:0044422	organelle part	2.63E-23	5.77E-21	1.48 (9133,2844,1042,479)
GO:0044464	cell part	2.18E-21	4.17E-19	1.18 (9133,6421,1107,919)
GO:0044444	cytoplasmic part	2.18E-17	3.71E-15	1.37 (9133,3121,1117,523)
GO:0005634	nucleus	4.93E-15	7.56E-13	1.44 (9133,2417,1028,392)
GO:0044428	nuclear part	2.40E-14	3.34E-12	1.53 (9133,1650,1135,313)
GO:0005829	cytosol	5.70E-11	7.29E-09	1.53 (9133,1201,1102,222)
GO:0032991	protein-containing complex	9.29E-11	1.10E-08	1.38 (9133,2090,1040,329)
GO:1902494	catalytic complex	1.25E-08	1.37E-06	1.78 (9133,482,1131,106)
GO:0005654	nucleoplasm	2.64E-08	2.70E-06	1.54 (9133,868,1132,166)

Table 5.16: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{-/-}$ SI compared to *C11orf*$^{+/-}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.16a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0002377	immunoglobulin production	7.67E-30	1.01E-25	22.22 (10086,62,205,28)
GO:0002440	production of molecular mediator of immune response	3.63E-28	2.39E-24	19.97 (10086,69,205,28)
GO:0006955	immune response	2.43E-17	1.07E-13	4.81 (10086,456,216,47)
GO:0002376	immune system process	6.88E-13	2.27E-09	3.93 (10086,753,133,39)
GO:0006958	complement activation, classical pathway	6.47E-12	1.70E-08	11.32 (10086,66,216,16)
GO:0006910	phagocytosis, recognition	1.56E-11	3.42E-08	11.93 (10086,59,215,15)
GO:0006956	complement activation	2.31E-11	4.35E-08	10.52 (10086,71,216,16)
GO:0072376	protein activation cascade	5.59E-11	9.20E-08	9.96 (10086,75,216,16)
GO:0006911	phagocytosis, engulfment	9.69E-11	1.42E-07	9.53 (10086,70,242,16)
GO:0099024	plasma membrane invagination	3.06E-10	4.03E-07	8.89 (10086,75,242,16)
GO:0050853	B cell receptor signaling pathway	3.50E-10	4.19E-07	9.77 (10086,72,215,15)
GO:0010324	membrane invagination	9.03E-10	9.90E-07	8.34 (10086,80,242,16)
GO:0006959	humoral immune response	1.03E-08	1.05E-05	3.23 (10086,111,927,33)
GO:0050871	positive regulation of B cell activation	1.15E-08	1.08E-05	7.73 (10086,91,215,15)
GO:0008037	cell recognition	1.72E-08	1.51E-05	6.95 (10086,108,215,16)
5.16b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	6.58E-12	2.42E-08	12.57 (10086,56,215,15)
GO:0003823	antigen binding	5.63E-11	1.04E-07	8.36 (10086,89,244,18)
GO:0004930	G-protein coupled receptor activity	6.81E-06	8.35E-03	4.49 (10086,151,223,15)

5.16c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0042571	immunoglobulin complex, circulating	3.60E-12	6.15E-09	13.03 (10086,54,215,15)
GO:0019814	immunoglobulin complex	8.55E-12	7.29E-09	12.35 (10086,57,215,15)
GO:0044459	plasma membrane part	4.53E-09	2.58E-06	1.55 (10086,1289,908,180)
GO:0009897	external side of plasma membrane	7.17E-09	3.06E-06	3.49 (10086,195,459,31)
GO:0044421	extracellular region part	5.20E-08	1.77E-05	2.58 (10086,795,216,44)
GO:0098552	side of membrane	2.20E-07	6.26E-05	2.91 (10086,240,476,33)
GO:0005576	extracellular region	3.38E-06	8.23E-04	1.55 (10086,699,1157,124)
GO:0097060	synaptic membrane	3.28E-05	7.00E-03	2.25 (10086,182,837,34)
GO:0044449	contractile fiber part	1.07E-04	1.82E-02	2.27 (10086,122,1130,31)
GO:0005887	integral component of plasma membrane	9.82E-05	1.86E-02	1.76 (10086,432,782,59)
GO:0031226	intrinsic component of plasma membrane	2.22E-04	3.44E-02	1.70 (10086,462,782,61)

Table 5.17: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/-}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.17a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0044237	cellular metabolic process	1.05E-12	1.37E-08	1.28 (10539,3679,1269,565)
GO:0043170	macromolecule metabolic process	6.58E-12	4.27E-08	1.31 (10539,2972,1269,470)
GO:0044260	cellular macromolecule metabolic process	4.24E-11	1.83E-07	1.34 (10539,2465,1269,399)
GO:0006807	nitrogen compound metabolic process	1.37E-10	4.43E-07	1.27 (10539,3388,1269,517)
GO:0048522	positive regulation of cellular process	1.83E-10	4.73E-07	1.34 (10539,2479,1203,380)
GO:0009987	cellular process	4.10E-10	8.86E-07	1.13 (10539,6743,1264,914)

GO:0034641	cellular nitrogen compound metabolic process	1.03E-09	1.67E-06	1.37 (10539,2004,1269,330)
GO:0071704	organic substance metabolic process	9.75E-10	1.80E-06	1.23 (10539,3893,1269,576)
GO:0090304	nucleic acid metabolic process	1.89E-09	2.45E-06	1.44 (10539,1486,1269,257)
GO:0008152	metabolic process	1.88E-09	2.70E-06	1.21 (10539,4179,1269,610)
GO:0044238	primary metabolic process	2.95E-09	3.47E-06	1.23 (10539,3700,1269,549)
GO:0006910	phagocytosis, recognition	7.59E-09	8.20E-06	100.95 (10539,58,9,5)
GO:0046483	heterocycle metabolic process	9.34E-09	9.31E-06	1.37 (10539,1805,1269,298)
GO:0006958	complement activation, classical pathway	1.18E-08	9.56E-06	92.94 (10539,63,9,5)
GO:0051338	regulation of transferase activity	1.12E-08	9.66E-06	1.86 (10539,437,1218,94)
5.17b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0005488	binding	1.12E-10	4.01E-07	1.15 (10539,6188,1251,846)
GO:0019899	enzyme binding	1.49E-09	2.67E-06	1.52 (10539,1153,1218,203)
GO:0034987	immunoglobulin receptor binding	6.78E-09	8.12E-06	102.72 (10539,57,9,5)
GO:0003823	antigen binding	6.02E-08	5.40E-05	68.88 (10539,85,9,5)
GO:0005515	protein binding	1.52E-07	1.09E-04	1.21 (10539,4298,1003,495)
GO:0003676	nucleic acid binding	7.24E-06	4.33E-03	1.35 (10539,1445,1179,219)
GO:0003723	RNA binding	9.96E-06	5.11E-03	1.68 (10539,441,1179,83)
GO:0008094	DNA-dependent ATPase activity	1.42E-05	6.38E-03	4.27 (10539,26,1233,13)
GO:0043167	ion binding	2.16E-05	8.62E-03	1.22 (10539,2674,1256,389)
GO:1901363	heterocyclic compound binding	4.21E-05	1.51E-02	1.23 (10539,2422,1269,358)
GO:0097159	organic cyclic compound binding	4.93E-05	1.61E-02	1.22 (10539,2472,1269,364)
GO:0003678	DNA helicase activity	1.64E-04	4.92E-02	9.13 (10539,16,433,6)

5.17c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0044424	intracellular part	7.27E-20	1.19E-16	1.20 (10539,6661,1259,952)
GO:0044464	cell part	6.28E-18	3.43E-15	1.15 (10539,7662,1259,1051)
GO:0043231	intracellular membrane-bounded organelle	4.48E-18	3.68E-15	1.28 (10539,4652,1269,719)
GO:0043227	membrane-bounded organelle	1.08E-17	4.42E-15	1.26 (10539,4978,1269,757)
GO:0043229	intracellular organelle	4.25E-17	1.39E-14	1.24 (10539,5393,1253,795)
GO:0043226	organelle	9.80E-17	2.68E-14	1.23 (10539,5483,1253,804)
GO:0044446	intracellular organelle part	2.91E-14	6.82E-12	1.33 (10539,3451,1258,548)
GO:0044422	organelle part	8.09E-14	1.66E-11	1.30 (10539,3583,1250,551)
GO:0005654	nucleoplasm	8.08E-13	1.47E-10	1.67 (10539,966,1269,194)
GO:0044444	cytoplasmic part	1.24E-12	2.04E-10	1.25 (10539,4041,1266,609)
GO:0044428	nuclear part	1.58E-12	2.36E-10	1.48 (10539,1857,1269,330)
GO:0032991	protein-containing complex	3.23E-12	4.41E-10	1.36 (10539,2588,1234,411)
GO:0005634	nucleus	1.22E-11	1.53E-09	1.32 (10539,2799,1269,446)
GO:0005737	cytoplasm	4.97E-11	5.83E-09	1.28 (10539,3220,1267,497)
GO:0005829	cytosol	9.68E-10	1.06E-07	1.50 (10539,1469,1053,220)

Table 5.18: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{-/-}$ LI compared to *C11orf*$^{+/-}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.18a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0006955	immune response	1.43E-24	9.10E-21	4.07 (9238,455,379,76)
GO:0050853	B cell receptor signaling pathway	1.19E-24	1.52E-20	11.51 (9238,76,338,32)

GO:0002440	production of molecular mediator of immune response	2.34E-22	7.44E-19	10.29 (9238,74,376,31)
GO:0002376	immune system process	2.26E-22	9.59E-19	3.12 (9238,749,380,96)
GO:0002377	immunoglobulin production	6.64E-22	1.69E-18	10.25 (9238,65,416,30)
GO:0002768	immune response-regulating cell surface receptor signaling pathway	1.31E-21	2.78E-18	6.98 (9238,124,427,40)
GO:0006910	phagocytosis, recognition	2.55E-21	4.64E-18	9.83 (9238,66,427,30)
GO:0006958	complement activation, classical pathway	9.12E-21	1.45E-17	7.27 (9238,72,618,35)
GO:0050851	antigen receptor-mediated signaling pathway	3.31E-20	4.22E-17	8.43 (9238,107,338,33)
GO:0002764	immune response-regulating signaling pathway	3.16E-20	4.46E-17	6.14 (9238,148,427,42)
GO:0002429	immune response-activating cell surface receptor signaling pathway	9.83E-20	1.14E-16	7.88 (9238,118,338,34)
GO:0006956	complement activation	1.65E-19	1.75E-16	6.79 (9238,77,618,35)
GO:0006911	phagocytosis, engulfment	1.75E-18	1.72E-15	16.26 (9238,76,157,21)
GO:0002757	immune response-activating signal transduction	2.84E-18	2.59E-15	5.98 (9238,141,427,39)
GO:0072376	protein activation cascade	6.07E-18	4.83E-15	6.23 (9238,84,618,35)
5.18b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	7.69E-23	2.74E-19	12.55 (9238,61,338,28)
GO:0003823	antigen binding	5.83E-19	1.04E-15	9.22 (9238,86,338,29)
GO:0005102	signaling receptor binding	7.88E-08	9.36E-05	1.55 (9238,818,1112,153)
GO:0022836	gated channel activity	1.35E-07	1.20E-04	2.39 (9238,161,1056,44)
GO:0022839	ion gated channel activity	2.24E-07	1.60E-04	2.38 (9238,158,1056,43)
GO:0005216	ion channel activity	1.16E-06	6.91E-04	2.09 (9238,203,1110,51)
GO:0005261	cation channel activity	1.50E-06	7.63E-04	2.28 (9238,157,1057,41)
GO:0022838	substrate-specific channel activity	2.95E-06	1.31E-03	2.04 (9238,208,1110,51)

GO:0002134	UTP binding	9.52E-06	2.42E-03	318.55 (9238,2,29,2)
GO:0015267	channel activity	6.91E-06	2.46E-03	1.97 (9238,220,1110,52)
GO:0032551	pyrimidine ribonucleoside binding	9.52E-06	2.61E-03	318.55 (9238,2,29,2)
GO:0022803	passive transmembrane transporter activity	6.91E-06	2.74E-03	1.97 (9238,220,1110,52)
GO:0008324	cation transmembrane transporter activity	8.46E-06	2.74E-03	1.87 (9238,278,1069,60)
GO:0032557	pyrimidine ribonucleotide binding	9.52E-06	2.83E-03	318.55 (9238,2,29,2)
GO:0015318	inorganic molecular entity transmembrane transporter activity	3.17E-05	7.54E-03	1.69 (9238,354,1110,72)
5.18c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	4.41E-25	7.23E-22	13.67 (9238,58,338,29)
GO:0042571	immunoglobulin complex, circulating	8.22E-23	6.74E-20	12.19 (9238,56,379,28)
GO:0009897	external side of plasma membrane	5.02E-18	2.74E-15	5.17 (9238,238,338,45)
GO:0098552	side of membrane	1.04E-15	4.24E-13	4.56 (9238,270,338,45)
GO:0044425	membrane part	1.48E-12	4.87E-10	1.31 (9238,3006,1118,476)
GO:0044421	extracellular region part	1.73E-10	4.06E-08	1.66 (9238,803,1108,160)
GO:0031224	intrinsic component of membrane	2.09E-10	4.28E-08	1.33 (9238,2370,1110,380)
GO:0016021	integral component of membrane	1.59E-10	4.35E-08	1.35 (9238,2283,1110,369)
GO:0044459	plasma membrane part	1.58E-08	2.88E-06	1.53 (9238,1283,845,179)
GO:0097458	neuron part	3.76E-07	6.16E-05	1.50 (9238,934,1061,161)
GO:0098793	presynapse	1.81E-06	2.69E-04	2.65 (9238,105,1061,32)
GO:0019815	B cell receptor complex	4.80E-06	6.56E-04	27.33 (9238,4,338,4)
GO:0042555	MCM complex	5.55E-06	7.00E-04	46.13 (9238,9,89,4)
GO:0005886	plasma membrane	8.30E-06	9.71E-04	1.28 (9238,1896,1118,294)
GO:0044456	synapse part	2.54E-05	2.77E-03	1.68 (9238,430,985,77)

Table 5.19: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{+/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.19a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0006955	immune response	6.93E-23	9.04E-19	13.85 (10593,510,39,26)
GO:0002377	immunoglobulin production	8.92E-20	5.82E-16	53.56 (10593,71,39,14)
GO:0002440	production of molecular mediator of immune response	5.04E-19	2.19E-15	48.13 (10593,79,39,14)
GO:0002376	immune system process	2.70E-17	8.79E-14	8.55 (10593,826,39,26)
GO:0006910	phagocytosis, recognition	5.83E-17	1.52E-13	43.87 (10593,73,43,13)
GO:0006958	complement activation, classical pathway	2.29E-16	4.98E-13	40.03 (10593,80,43,13)
GO:0050853	B cell receptor signaling pathway	3.96E-16	6.45E-13	38.58 (10593,83,43,13)
GO:0006956	complement activation	4.73E-16	6.85E-13	38.13 (10593,84,43,13)
GO:0006911	phagocytosis, engulfment	3.96E-16	7.37E-13	38.58 (10593,83,43,13)
GO:0072376	protein activation cascade	9.40E-16	1.23E-12	36.39 (10593,88,43,13)
GO:0099024	plasma membrane invagination	1.31E-15	1.55E-12	35.58 (10593,90,43,13)
GO:0008037	cell recognition	1.88E-15	2.04E-12	28.74 (10593,120,43,14)
GO:0010324	membrane invagination	2.48E-15	2.49E-12	34.07 (10593,94,43,13)
GO:0050871	positive regulation of B cell activation	1.64E-14	1.53E-11	29.93 (10593,107,43,13)
GO:0002429	immune response-activating cell surface receptor signaling pathway	3.81E-14	3.31E-11	25.22 (10593,127,43,13)
5.19b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	2.50E-17	9.18E-14	46.41 (10593,69,43,13)
GO:0003823	antigen binding	4.21E-16	7.75E-13	31.64 (10593,109,43,14)
GO:0019899	enzyme binding	1.01E-05	1.23E-02	1.39 (10593,1232,1157,187)

GO:0016740	transferase activity	3.36E-05	3.09E-02	1.37 (10593,1132,1274,186)
GO:0019900	kinase binding	7.07E-05	4.34E-02	1.75 (10593,436,876,63)
GO:0005515	protein binding	6.39E-05	4.70E-02	1.14 (10593,4616,1275,632)
5.19c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	3.10E-17	2.63E-14	45.75 (10593,70,43,13)
GO:0042571	immunoglobulin complex, circulating	1.60E-17	2.71E-14	47.80 (10593,67,43,13)
GO:0009897	external side of plasma membrane	6.94E-13	3.92E-10	15.89 (10593,217,43,14)
GO:0098552	side of membrane	9.97E-12	4.22E-09	13.16 (10593,262,43,14)
GO:0044464	cell part	6.81E-07	2.31E-04	1.08 (10593,8048,1270,1040)
GO:0044451	nucleoplasm part	9.30E-07	2.63E-04	1.66 (10593,533,1248,104)
GO:0016604	nuclear body	7.86E-06	1.90E-03	1.86 (10593,334,1056,62)
GO:0016020	membrane	9.08E-06	1.92E-03	1.18 (10593,3965,1149,508)
GO:0044428	nuclear part	2.10E-05	3.96E-03	1.26 (10593,2077,1275,315)
GO:0005829	cytosol	2.80E-05	4.74E-03	1.29 (10593,1699,1276,264)
GO:0043226	organelle	3.68E-05	5.67E-03	1.11 (10593,5964,1276,796)
GO:0043227	membrane-bounded organelle	6.35E-05	8.97E-03	1.12 (10593,5423,1276,730)
GO:0044424	intracellular part	8.29E-05	1.08E-02	1.08 (10593,7084,1276,923)
GO:0043229	intracellular organelle	9.43E-05	1.14E-02	1.11 (10593,5865,1276,781)
GO:0043231	intracellular membrane-bounded organelle	1.18E-04	1.33E-02	1.12 (10593,5084,1276,687)

Table 5.20: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{+/-}$ SI compared to *C11orf*$^{+/+}$ SI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.20a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0072376	protein activation cascade	6.55E-19	8.25E-15	32.30 (8706,77,56,16)
GO:0006910	phagocytosis, recognition	3.01E-18	1.90E-14	21.53 (8706,56,130,18)
GO:0006956	complement activation	3.69E-17	1.55E-13	17.13 (8706,68,142,19)
GO:0006958	complement activation, classical pathway	6.91E-17	2.18E-13	18.39 (8706,60,142,18)
GO:0006955	immune response	1.68E-16	4.23E-13	15.91 (8706,443,21,17)
GO:0002376	immune system process	1.03E-15	2.17E-12	10.05 (8706,722,24,20)
GO:0050853	B cell receptor signaling pathway	2.08E-15	3.75E-12	17.52 (8706,65,130,17)
GO:0006911	phagocytosis, engulfment	3.78E-15	5.95E-12	16.99 (8706,67,130,17)
GO:0099024	plasma membrane invagination	6.71E-15	9.40E-12	16.50 (8706,69,130,17)
GO:0010324	membrane invagination	1.17E-14	1.47E-11	16.03 (8706,71,130,17)
GO:0006959	humoral immune response	2.40E-14	2.75E-11	10.85 (8706,113,142,20)
GO:0050851	antigen receptor-mediated signaling pathway	3.99E-14	4.19E-11	12.69 (8706,95,130,18)
GO:0002377	immunoglobulin production	4.43E-14	4.30E-11	5.56 (8706,59,769,29)
GO:0002440	production of molecular mediator of immune response	6.20E-14	5.58E-11	5.01 (8706,70,769,31)
GO:0008037	cell recognition	1.75E-13	1.38E-10	11.70 (8706,103,130,18)
5.20b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	1.36E-18	4.61E-15	22.32 (8706,54,130,18)
GO:0003823	antigen binding	5.05E-15	8.57E-12	16.74 (8706,68,130,17)
GO:0038023	signaling receptor activity	2.12E-11	2.40E-08	1.80 (8706,779,852,137)

GO:0060089	molecular transducer activity	2.85E-11	2.42E-08	1.79 (8706,782,852,137)
GO:0004888	transmembrane signaling receptor activity	8.76E-11	5.95E-08	1.85 (8706,703,803,120)
GO:0004984	olfactory receptor activity	8.44E-08	4.78E-05	2.69 (8706,283,458,40)
GO:0005102	signaling receptor binding	1.58E-05	7.68E-03	5.48 (8706,728,24,11)
5.20c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	1.02E-17	7.89E-15	22.77 (8706,50,130,17)
GO:0042571	immunoglobulin complex, circulating	6.66E-18	1.03E-14	23.23 (8706,49,130,17)
GO:0009897	external side of plasma membrane	6.52E-11	3.37E-08	2.55 (8706,238,847,59)
GO:0098552	side of membrane	2.38E-10	9.24E-08	6.09 (8706,259,116,21)
GO:0031224	intrinsic component of membrane	1.90E-05	4.21E-03	1.24 (8706,2292,1015,332)
GO:0044459	plasma membrane part	1.80E-05	4.64E-03	1.36 (8706,1233,1018,196)
GO:0044425	membrane part	1.60E-05	4.98E-03	1.21 (8706,2834,1019,401)
GO:0016021	integral component of membrane	1.03E-04	1.99E-02	1.23 (8706,2229,1029,323)
GO:0044421	extracellular region part	1.34E-04	2.30E-02	4.96 (8706,731,24,10)

Table 5.21: Gene ontology enrichment analysis of genes upregulated in *C11orf*$^{+/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.21a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0006955	immune response	1.66E-17	2.06E-13	15.24 (9104,323,37,20)
GO:0002377	immunoglobulin production	8.10E-16	5.03E-12	76.63 (9104,44,27,10)
GO:0002440	production of molecular mediator of immune response	4.78E-15	1.98E-11	66.11 (9104,51,27,10)
GO:0002376	immune system process	2.26E-14	7.03E-11	9.13 (9104,539,37,20)

GO:0006959	humoral immune response	1.26E-11	3.12E-08	14.26 (9104,109,82,14)
GO:0050853	B cell receptor signaling pathway	6.92E-11	1.23E-07	45.18 (9104,52,31,8)
GO:0006958	complement activation, classical pathway	8.23E-11	1.28E-07	44.33 (9104,53,31,8)
GO:0006910	phagocytosis, recognition	6.92E-11	1.43E-07	45.18 (9104,52,31,8)
GO:0006911	phagocytosis, engulfment	1.65E-10	2.05E-07	41.22 (9104,57,31,8)
GO:0006956	complement activation	1.65E-10	2.28E-07	41.22 (9104,57,31,8)
GO:0099024	plasma membrane invagination	3.28E-10	3.39E-07	37.89 (9104,62,31,8)
GO:0045087	innate immune response	3.01E-10	3.40E-07	9.46 (9104,176,82,15)
GO:0010324	membrane invagination	4.38E-10	4.18E-07	36.71 (9104,64,31,8)
GO:0072376	protein activation cascade	4.94E-10	4.38E-07	36.14 (9104,65,31,8)
GO:0042742	defense response to bacterium	6.13E-10	5.07E-07	20.34 (9104,121,37,10)
5.21b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0003823	antigen binding	3.28E-12	1.17E-08	44.80 (9104,59,31,9)
GO:0034987	immunoglobulin receptor binding	6.92E-11	1.23E-07	45.18 (9104,52,31,8)
GO:0043167	ion binding	1.18E-06	1.40E-03	1.27 (9104,2393,1021,341)
GO:0004527	exonuclease activity	8.42E-06	7.49E-03	4.13 (9104,38,871,15)
5.21c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0019814	immunoglobulin complex	3.52E-11	2.93E-08	48.95 (9104,48,31,8)
GO:0042571	immunoglobulin complex, circulating	2.91E-11	4.84E-08	49.99 (9104,47,31,8)
GO:0044464	cell part	7.66E-08	4.26E-05	1.09 (9104,7006,1019,855)
GO:0009897	external side of plasma membrane	1.62E-07	6.73E-05	15.87 (9104,148,31,8)
GO:0098552	side of membrane	6.15E-07	2.05E-04	13.43 (9104,175,31,8)
GO:0044424	intracellular part	3.04E-06	8.45E-04	1.10 (9104,6256,1029,778)

GO:0016020	membrane	3.89E-06	9.27E-04	1.21 (9104,3257,1007,435)
GO:0043229	intracellular organelle	9.41E-05	1.96E-02	1.11 (9104,5242,1029,659)
GO:0043226	organelle	1.16E-04	2.15E-02	1.11 (9104,5340,1029,669)
GO:0005911	cell-cell junction	1.37E-04	2.29E-02	1.96 (9104,174,1041,39)
GO:0005886	plasma membrane	1.99E-04	3.01E-02	1.29 (9104,1522,1017,219)
GO:0043226	organelle	1.16E-04	2.15E-02	1.11 (9104,5340,1029,669)

Table 5.22: Gene ontology enrichment analysis of genes downregulated in *C11orf*$^{+/-}$ LI compared to *C11orf*$^{+/+}$ LI. Only results with FDR q-value lower than 0.05 are included; only the top 15 results are shown if more than that were significant, with full results in appendix 3.

5.22a) Cellular process				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0006955	immune response	6.91E-27	9.09E-23	2.74 (10641,618,868,138)
GO:0006910	phagocytosis, recognition	5.48E-23	2.40E-19	13.47 (10641,76,291,28)
GO:0002376	immune system process	5.25E-23	3.45E-19	2.06 (10641,1005,1042,203)
GO:0006958	complement activation, classical pathway	1.68E-22	5.53E-19	12.33 (10641,86,291,29)
GO:0002377	immunoglobulin production	2.48E-21	6.52E-18	32.74 (10641,78,75,18)
GO:0006956	complement activation	3.34E-21	7.32E-18	11.28 (10641,94,291,29)
GO:0050871	positive regulation of B cell activation	5.94E-21	1.12E-17	6.91 (10641,115,522,39)
GO:0072376	protein activation cascade	1.32E-20	2.17E-17	10.82 (10641,98,291,29)
GO:0006911	phagocytosis, engulfment	1.83E-20	2.67E-17	18.57 (10641,92,137,22)
GO:0050853	B cell receptor signaling pathway	2.62E-20	3.45E-17	7.58 (10641,95,517,35)
GO:0002440	production of molecular mediator of immune response	4.79E-20	5.73E-17	28.38 (10641,90,75,18)
GO:0099024	plasma membrane invagination	5.28E-20	5.79E-17	17.80 (10641,96,137,22)

GO:0010324	membrane invagination	1.45E-19	1.47E-16	17.09 (10641,100,137,22)
GO:0008037	cell recognition	1.50E-18	1.41E-15	8.50 (10641,129,301,31)
GO:0050864	regulation of B cell activation	4.63E-18	4.06E-15	5.74 (10641,142,522,40)
5.22b) Molecular function				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0034987	immunoglobulin receptor binding	3.28E-24	1.14E-20	14.63 (10641,70,291,28)
GO:0003823	antigen binding	9.25E-24	1.61E-20	17.71 (10641,114,137,26)
GO:0038023	signaling receptor activity	1.33E-07	1.55E-04	1.51 (10641,1002,1188,169)
GO:0060089	molecular transducer activity	1.77E-07	1.55E-04	1.50 (10641,1006,1188,169)
GO:0004930	G-protein coupled receptor activity	4.97E-06	3.47E-03	1.86 (10641,300,1201,63)
GO:0004888	transmembrane signaling receptor activity	2.47E-05	1.44E-02	1.44 (10641,903,1186,145)
GO:0030246	carbohydrate binding	2.97E-05	1.48E-02	2.85 (10641,152,565,23)
GO:0005102	signaling receptor binding	8.07E-05	3.52E-02	5.77 (10641,922,18,9)
5.22c) Cellular component				
GO term	**Description**	**P-value**	**FDR q-value**	**Enrichment (N, B, n, b)**
GO:0042571	immunoglobulin complex, circulating	1.19E-24	1.91E-21	15.06 (10641,68,291,28)
GO:0019814	immunoglobulin complex	5.36E-24	4.29E-21	14.42 (10641,71,291,28)
GO:0009897	external side of plasma membrane	1.12E-20	5.98E-18	2.83 (10641,317,1173,99)
GO:0098552	side of membrane	3.69E-18	1.47E-15	2.60 (10641,356,1173,102)
GO:0044459	plasma membrane part	1.39E-13	4.46E-11	1.56 (10641,1551,1258,286)
GO:0044425	membrane part	2.88E-10	7.69E-08	1.28 (10641,3565,1051,452)
GO:0031224	intrinsic component of membrane	1.75E-05	3.99E-03	1.24 (10641,2854,1051,349)
GO:0031226	intrinsic component of plasma membrane	3.38E-05	6.76E-03	1.50 (10641,643,1266,115)
GO:0044421	extracellular region part	5.46E-05	7.94E-03	6.09 (10641,874,18,9)
GO:0016021	integral component of membrane	5.46E-05	8.74E-03	1.23 (10641,2756,1075,342)

GO:0005887	integral component of plasma membrane	5.28E-05	9.39E-03	1.51 (10641,601,1266,108)

5.6.2.3 Wnt signalling pathway

I looked at several key genes from the Wnt signalling pathway, as it is highly significant in CRC. Additionally, the altered expression levels of the *C11orf* genes between WT and *Apc*$^{Min/+}$ tissue seen with the RNAscope (figure 4.7) indicates an interaction between the Wnt and *C11orf* pathways. This pathway is extensive with many associated genes; as it was not practical to assess every one, I have chosen a subset of important proteins, or representative members of important families, to illustrate overall changes in the Wnt pathway.

There are 19 known members of the Wnt family of secreted ligands in the mouse (Nusse, 2018). Wnt3 is expressed in the colon from Paneth cells and is elevated in CRC tumours (Farin et al., 2012; Voloshanenko et al., 2013). *Wnt3* shows no difference in expression between the *C11orf* homozygous knockouts and WT, and although there does appear to be a reduction in the heterozygotes, it was not significant (figure 5.22a).

There are 13 known members of the Frizzled family of receptors in the mouse (Nusse, 2018). Of these, Fzd7 is the one predominantly expressed in CRC cell lines and may be a CRC therapeutic target (Ueno et al., 2008; Vincan et al., 2007). There is no difference in *Fzd7* expression between genotypes (figure 5.22b).

β-catenin, discussed above, is a key transcription factor which stimulates expression of Wnt-responsive genes. There is no difference in *Ctnnb1* expression between genotypes (figure 5.22c). This is in line with the β-catenin IHC staining (see figure 5.14).

As previously described, Apc is frequently mutated in CRC (Brannon et al., 2014). Apc forms part of the complex which targets β-catenin for degradation and is inhibited when Wnt signalling is activated, allowing β-catenin accumulation (Munemitsu et al., 1995). There does appear to be a reduction in *Apc* expression in both heterozygous and homozygous *C11orf* knockouts in the DLI but not in any other tissue section, and this was not a significant change – although the DLI data was pooled with PLI and R/A data for the differential analysis, which may have masked any change in only one section of the LI (figure 5.22d).

In addition to its function as a transcription factor in the Wnt signalling pathway, β-catenin also plays a role in cell adhesion junctions, where it binds to E-cadherin (Drees et al., 2005). E-cadherin and Apc compete for β-catenin (Hülsken et al., 1994) and loss of E-cadherin promotes growth, invasion and drug resistance in CRC (Chen et al., 2012). There is no difference in *Cdh1* expression between genotypes (figure 5.22e).

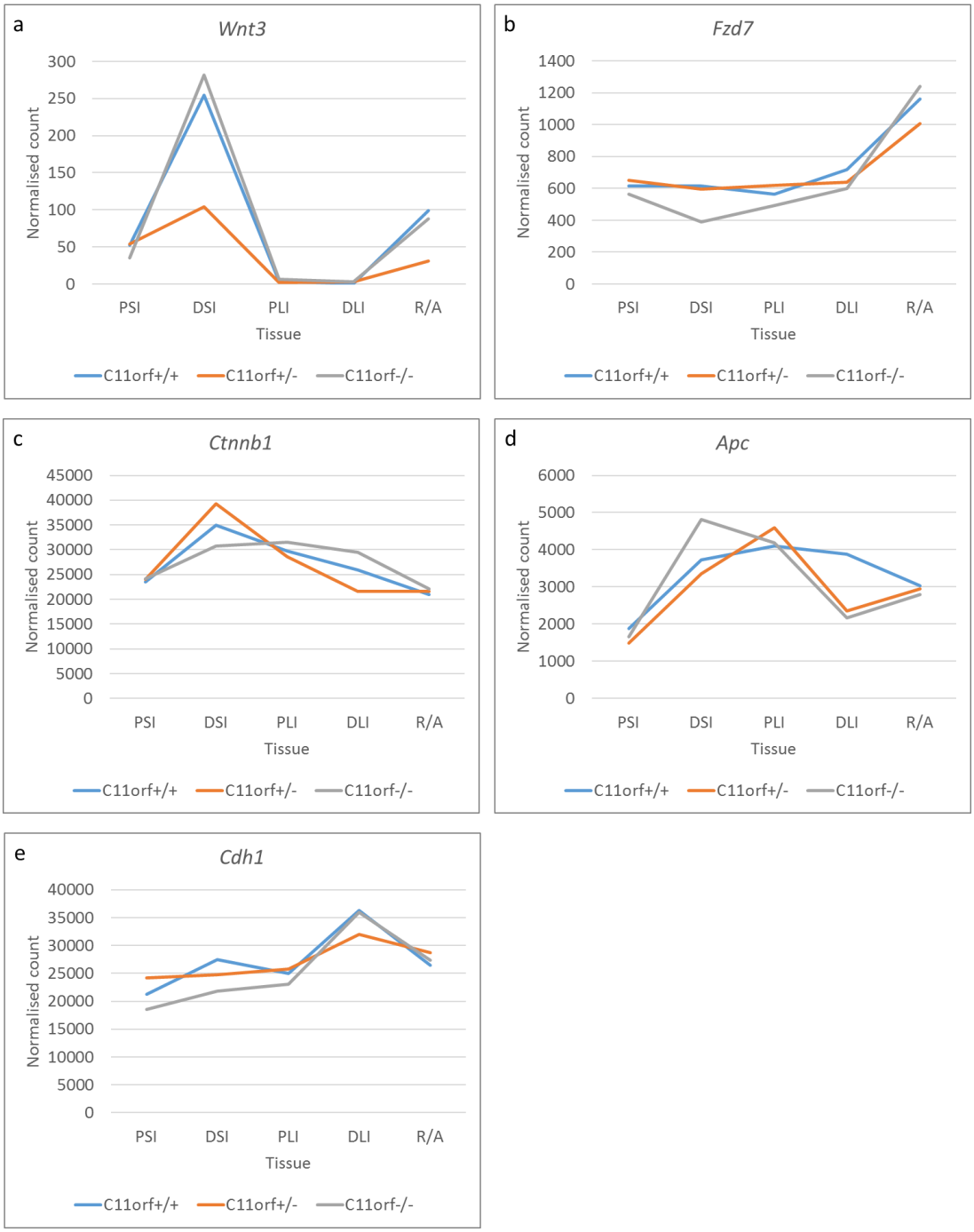


Figure 5.22: mRNA-seq expression data for Wnt signalling pathway components (a) *Wnt3*, (b) *Fzd7*, (c) *Ctnnb1*, (d) *Apc* and (e) *Cdh1*. Normalised counts for each genotype (*C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$) shown for each tissue analysed (proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and rectum/anus).

5.6.2.4 Goblet cell markers

Following the IHC results showing reduced mucin and Clca1 staining, I looked at the expression of *Muc2*, *Clca1*, *Fcgbp*, *Agr2* and *Zg16*, which are all key markers for goblet cell activity (Birchenough et al., 2015).

Muc2 is a mucin that forms the basis of intestinal mucus, the protective layer shielding the intestinal epithelium from damage, both by acting as a lubricant to aid the passage of material, and by providing a barrier against microbes. Deficiency in Muc2 has been shown to cause colitis (Van der Sluis et al., 2006) and CRC (Velcich et al., 2002). There is no difference in *Muc2* expression between genotypes (figure 5.23a).

Despite the IHC results showing reduced staining of Clca1, the chloride ion channel regulator, there is no difference in *Clca1* expression between genotypes (figure 5.23b).

Fcgbp is a glycoprotein which binds Muc2 to aid in stabilisation of the mucus layer through crosslinking; it also binds to the Fc part of IgG antibodies (Johansson et al., 2009). Fcgbp has been shown to regulate EMT in gallbladder cancer (Xiong et al., 2014). There appears to be a reduction in *Fcgbp* expression in heterozygous and homozygous *C11orf* mutants in the DLI (figure 5.23c). The gene is not a significant hit but the effect may be lost due to pooling of the DLI data with the PLI and R/A data, where there is no difference in expression.

Agr2 is a protein disulphide isomerase present in the ER to aid in the folding of secreted proteins; goblet cells have extensive ER due to the high levels of proteins secreted and Agr2 is essential for correct Muc2 production (Park et al., 2009; Zhao et al., 2010). Agr2 has been linked to IBD and CRC (Kim et al., 2011; Zheng et al., 2006). There is no difference in *Agr2* expression between genotypes (figure 5.23d).

Zg16 is a lectin-like component of mucus which binds Gram-positive bacteria to reduce their interaction with the intestinal wall (Bergström et al., 2016). Loss is associated with increased stemness and CRC development (Hasnain et al., 2013). There appears to be an increase in *Zg16* expression in heterozygous *C11orf* knockouts in the DLI, but *Zg16* was not listed as significantly altered (figure 5.23e).

Figure 5.23: mRNA-seq expression data for goblet cell markers (a) *Muc2*, (b) *Clca1*, (c) *Fcgbp*, (d) *Agr2* and (e) *Zg16*. Normalised counts for each genotype (*C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$) shown for each tissue analysed (proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and rectum/anus).

5.6.2.5 Tuft cell markers

As discussed in chapter 4, there are secretory cell types beyond goblet cells. I therefore looked at the tuft cell markers *Dclk1*, *Trpm5*, *Gnat3*, *Chat* and *Gfi1b* (Von Moltke et al., 2016).

Dclk1 is a kinase which binds microtubules and is involved in directed cargo transport (Lipka et al., 2016). In addition to being a tuft cell marker it can be used to distinguish normal stem cells (Dclk1-negative) from tumour stem cells (Dclk1-positive) in the intestine (Nakanishi et al., 2013). Dclk1-expressing cells are also required for intestinal regeneration and may function in maintaining the stem cell niche (Middelhoff et al., 2017) There is a reduction in *Dclk1* expression between *C11orf* WT and *C11orf* heterozygous knockouts, with a further reduction in the homozygous knockouts, although it was not a significantly altered gene (figure 5.24a).

Trpm5 is a cation channel activated by Ca^{2+} that is important in taste transduction (Liu and Liman, 2003). As with *Dclk1*, *Trpm5* shows reduced expression in both the heterozygous and homozygous *C11orf* knockouts, with a greater effect in the homozygotes (figure 5.24b). *Trpm5* is also a significant hit in the differential expression analysis between $C11orf^{-/-}$ and $C11orf^{+/+}$ in the LI, with a $\log_2$ fold change of -3.52 and adjusted p-value 3.45E-4.

Another tuft cell marker with a role in taste is Gnat3, a subunit of the G protein gustducin; loss of Gnat3 causes male sterility (Mosinger et al., 2013). *C11orf* homozygous knockouts show total loss of *Gnat3* expression, with a partial loss in heterozygotes in the DSI, PLI and DLI (figure 5.24c). It was not significant in the differential analysis, possibly because these samples were pooled with PSI and R/A samples, where the WT also has minimal expression so the fold change is very low. This may be involved in the infertility seen in the $C11orf^{-/-}$ males.

Choline acetyltransferase synthesises the neurotransmitter acetylcholine in cholinergic neurons. Acetylcholine is released by human CRC cells and mediates cell proliferation stimulation (Cheng et al., 2008). While *Chat* was not considered significantly altered, it does show reduction in proportion with *C11orf* reduction (figure 5.24d).

Gfi1b is a zinc-finger protein that is a key regulator of haematopoiesis, specifically being essential for megakaryocytic and erythroid development, with *Gfi1b* null mice dying during embryonic development due to lack of erythrocytes (Saleque et al., 2002). It is overexpressed

in leukaemia (Vassen et al., 2009). *C11orf*$^{-/-}$ and *C11orf*$^{+/-}$ mice show lower *Gfi1b* expression than *C11orf*$^{+/+}$, although it was not a significant hit (figure 5.24e).

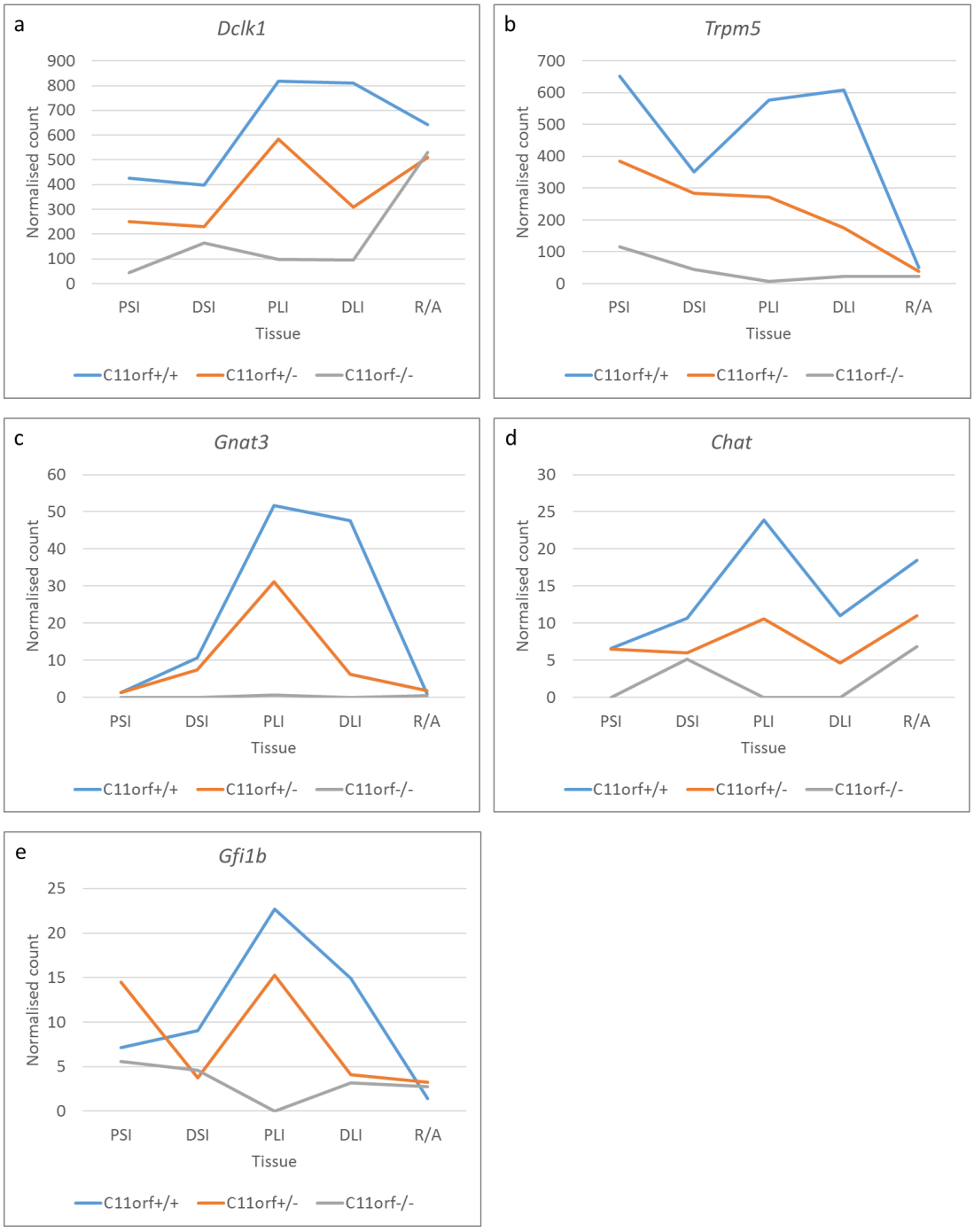


Figure 5.24: mRNA-seq expression data for tuft cell markers (a) *Dclk1*, (b) *Trpm5*, (c) *Gnat3*, (d) *Chat* and (e) *Gfi1b*. Normalised counts for each genotype (*C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$) shown for each tissue analysed (proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and rectum/anus).

5.6.2.6 Enteroendocrine cell markers

In addition to goblet cells and tuft cells, enteroendocrine cells are also a secretory cell type found in the intestine. *Cldn4*, *Chga*, *Ffar1* and *Ffar4* are enteroendocrine cell markers (Nagatake et al., 2014).

Claudin-4 is an integral membrane protein which is a component of tight junction strands and has been associated with multiple cancers, including breast, ovarian, prostate and pancreatic (Morin, 2005). There is no clear pattern of change in *Cldn4* expression between genotypes (figure 5.25a).

Chromogranin A plays roles in secretory granule biogenesis and calcium homeostasis, in addition to being the precursor to a number of other secretory peptides (D'amico et al., 2014). There is no difference in *Chga* expression between genotypes (figure 5.25b).

Free fatty acid receptors 1 and 4, also known as GPR40 and GPR120, are nutrient sensors involved in energy homeostasis regulation via secretion of hormones such as insulin, GLP-1 and CCK (Kaemmerer et al., 2010). There is no difference in either *Ffar1* or *Ffar4* expression between genotypes (figure 5.25c, 5.25d).

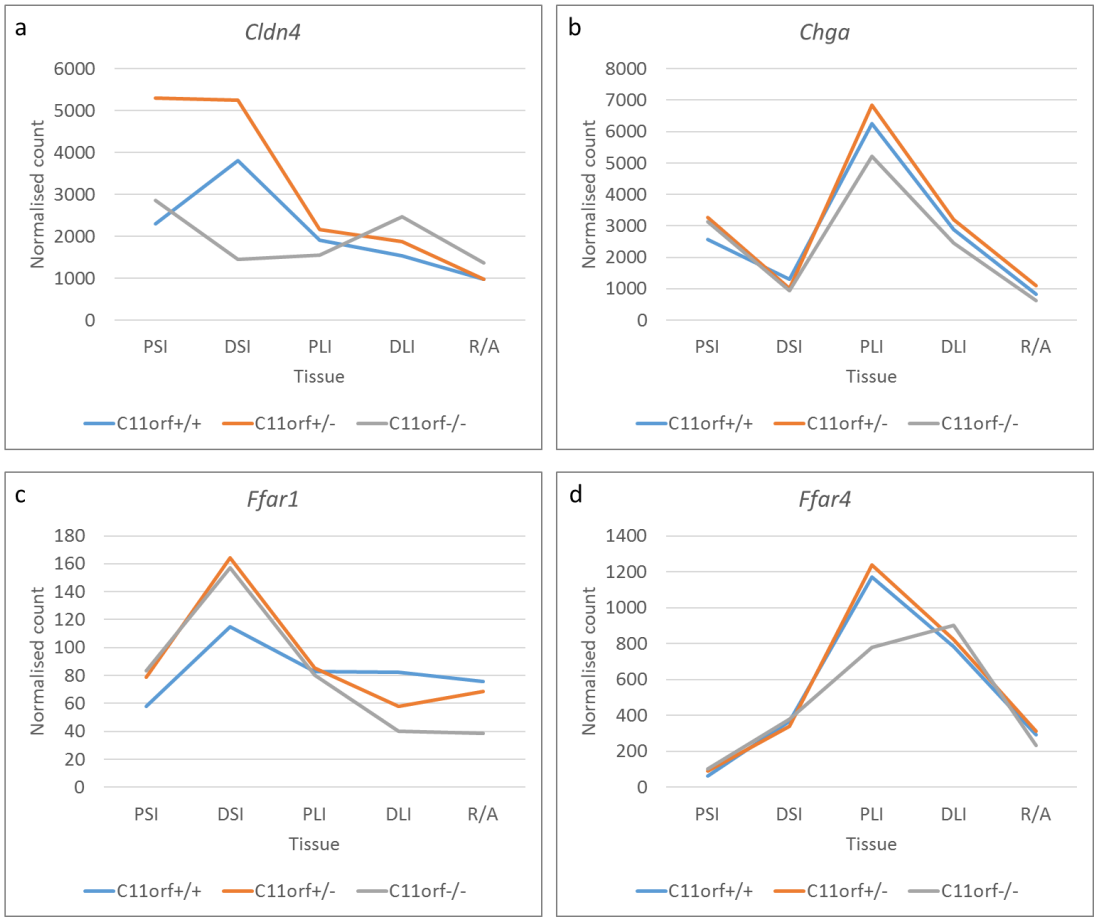


Figure 5.25: mRNA-seq expression data for enteroendocrine cell markers (a) *Cldn4*, (b) *Chga*, (c) *Ffar1* and (d) *Ffar4*. Normalised counts for each genotype (*C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$) shown for each tissue analysed (proximal small intestine, distal small intestine, proximal large intestine, distal large intestine, and rectum/anus).

5.7 Haematological analysis

5.7.1 Methods

I collected fresh blood via cardiac puncture from mice culled by CO_2 asphyxiation, which was stored on ice with EDTA anticoagulant. Full blood counts were performed by the University of Edinburgh Easter Bush Pathology Laboratory. Initial blood samples were all from 14-month old males (three wild type, three heterozygous knockouts and two homozygous knockouts), with further samples taken from two-month old males (one per genotype).

5.7.2 Results

The blood counts at 14 months of age show many significant differences in the blood of *C11orf*$^{-/-}$ mice compared to *C11orf*$^{+/-}$ and *C11orf*$^{+/+}$ mice. The homozygotes have greatly reduced total white blood cells (WBCs), with almost a seven-fold reduction from the heterozygotes (figure 5.26a). The differential counts of the WBC subtypes show the relative proportions are broadly conserved, although heterozygous and homozygous knockouts show decreased neutrophils and increased lymphocytes compared to wild types (figure 5.26b).

C11of$^{-/-}$ mice also have severely decreased red blood cells (RBCs) and somewhat lower mean corpuscular haemoglobin concentration, giving greatly reduced overall haemoglobin (figure 5.27). Their packed cell volume is low, but there is much variation in their mean corpuscular volume, with elevated red blood cell distribution width. Preliminary results from 7-week old males suggest the homozygotes are already deviating from the wild type at this age, although thus far the data is based on only a single mouse per genotype. Blood counts taken by the IMPC do not show any significant differences in the *C11orf93* knockout line compared to WT. However, IMPC blood tests are only performed at 16 weeks; our data shows that the *C11orf*$^{-/-}$ blood phenotype becomes more pronounced with age, so while at 16 weeks there may not be enough of a difference to be significant, the *C11orf93* knockout line may show a phenotype at a later age.

During general tissue collection from the colony I also observed that approximately one third of both homozygous and heterozygous knockouts had large blood clots present on their lungs when I opened the chest cavity, which were not found in wild type animals. Histological examination showed that the blood did not originate from the lungs, which appeared

healthy, and the blood was simply adhering to the surface of the lungs as it clotted (figure 5.28). This is only seen when the mice are culled by cervical dislocation, and not in those culled by CO_2 asphyxiation. Animals that are culled by asphyxiation and then undergo cervical dislocation immediately after death have this blood present at similar rates (table 5.23).

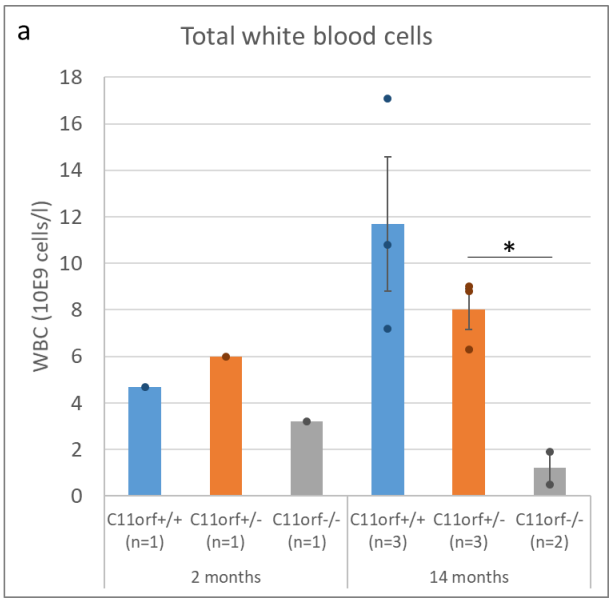
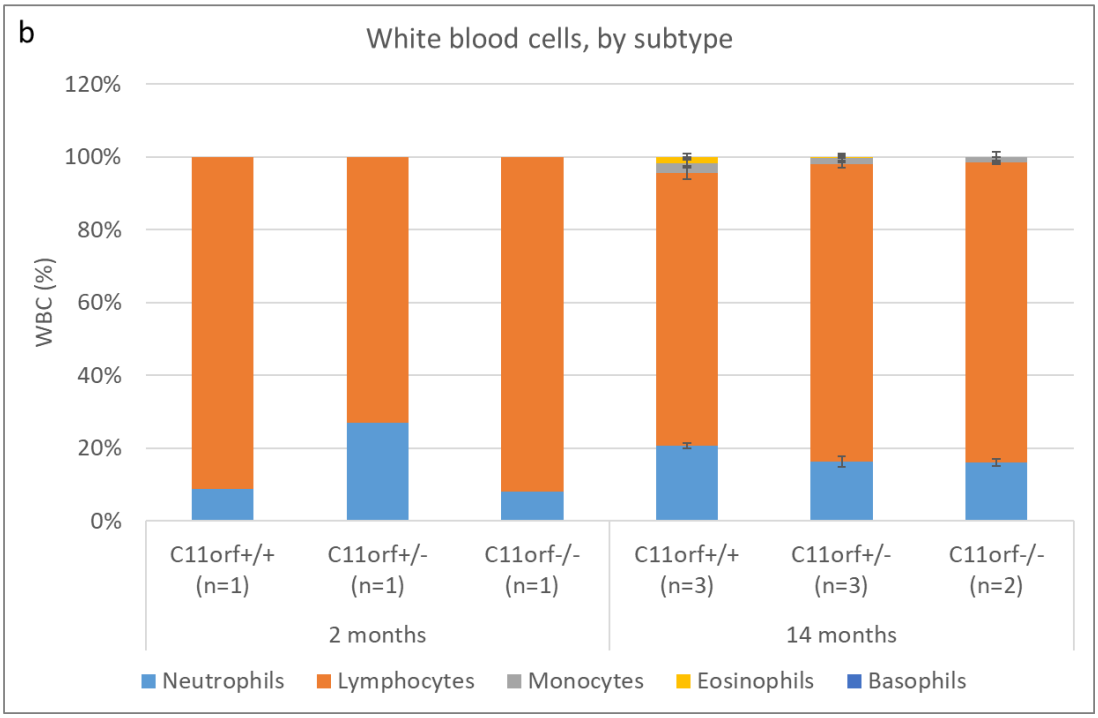


Figure 5.26: Analysis of blood samples taken from males. Bars represent the mean, with each animal displayed as an individual datapoint. (a) Total white blood cells; (b) proportion of white blood cell subtypes. Data was not normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test. Significant results are indicated by stars.



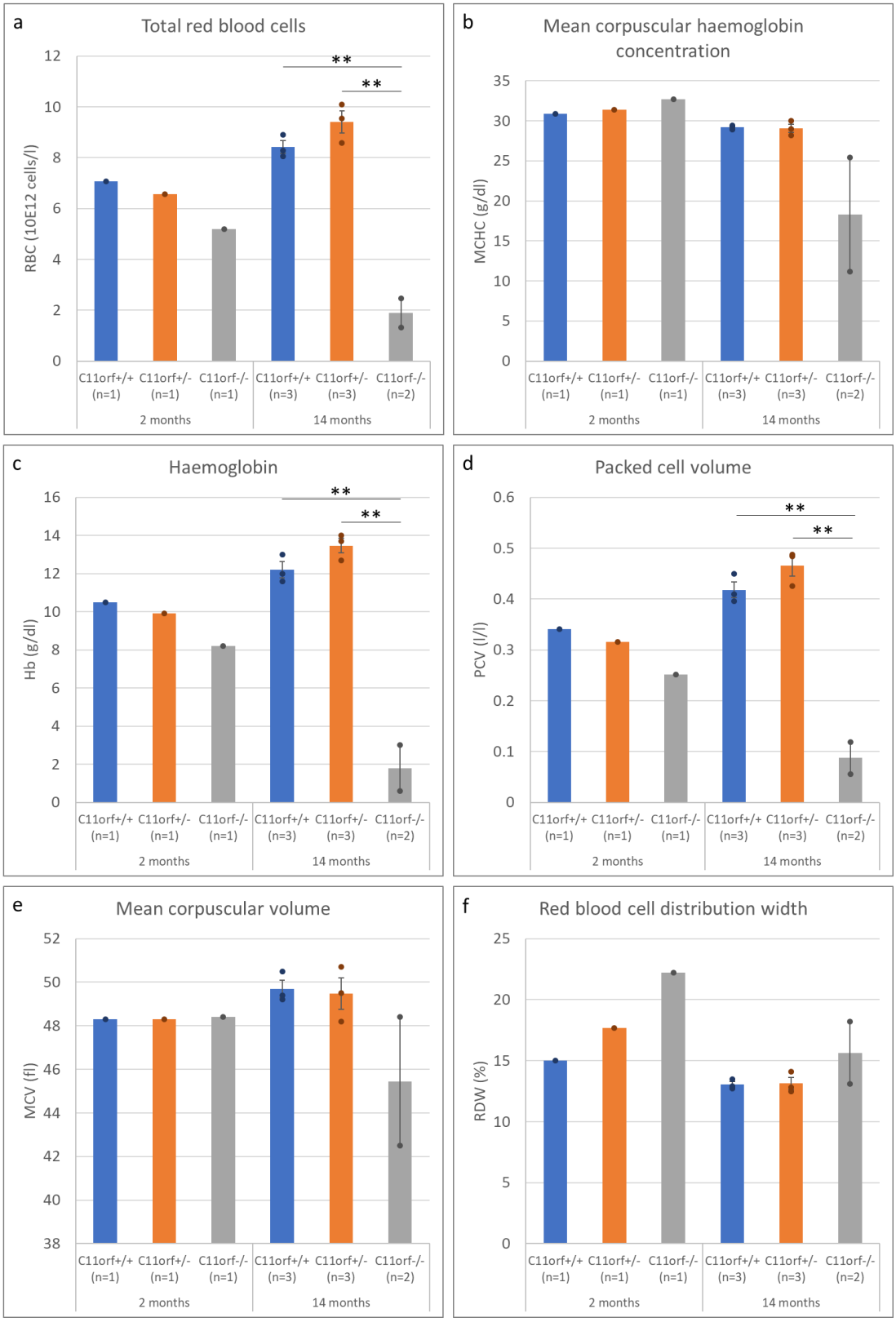


Figure 5.27: Analysis of blood samples taken from males. Bars represent the mean, with each animal displayed as an individual datapoint. (a) Total red blood cell count; (b) mean corpuscular haemoglobin concentration; (c) total haemoglobin concentration; (d) packed cell

volume, which is predominantly comprised of RBCs; (e) mean corpuscular volume; (f) red blood cell distribution width, indicating the variation in RBC width. Data was not normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test. Significant results are indicated by stars.

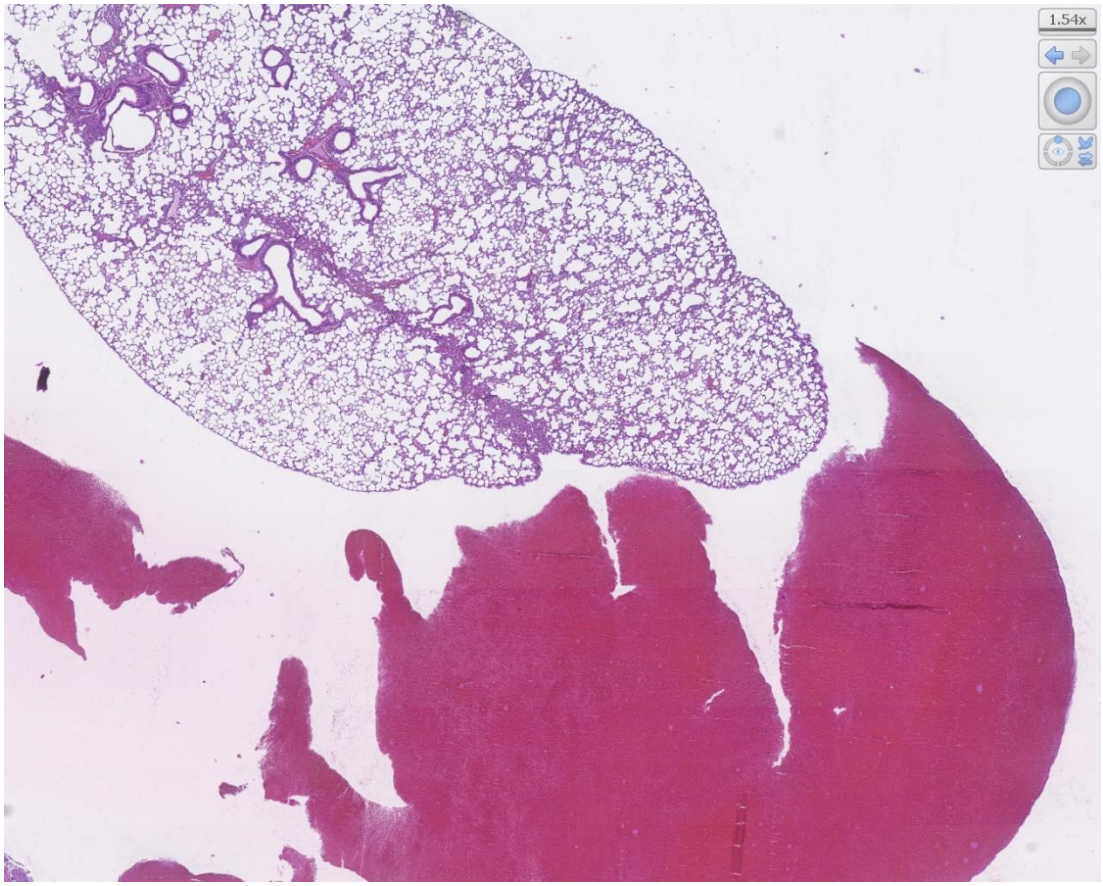


Figure 5.28: H&E of a lung and chest blood clot from a heterozygous knockout mouse, characteristic of the blood in the chest cavity observed in the colony at necropsy. The blood clot was adhered onto the lung at removal but no blood is present within the lung itself and the clot detached cleanly following fixation, indicating that the blood did not originate from the lungs.

Cull method	**Blood present in thoracic cavity**
Cervical dislocation (n=55)	31%
CO_2 asphyxiation (n=11)	0%
CO_2 asphyxiation + cervical dislocation (n=9)	33%

Table 5.23: Cause of death of animals carrying either one or two *C11orf* null mutations, and the percentage with blood present in the thoracic cavity during necropsy.

5.8 Discussion

I have successfully created a knockout mouse model of the locus containing *C11orf53*, *C11orf92* and *C11orf93*, as confirmed by sequencing and expression data. The animals show a clear phenotype of developmental issues, alongside altered intestinal homeostasis, haematopoietic system dysfunction and expression analysis indicating an immune system disruption.

The most striking phenotype in the triple gene knockout line is the reduction in homozygote frequency at birth and homozygote survival following weaning, both of which show some developmental problems in the young animals that do not affect adult animals if they survive to maturity. Those animals that survive to one month then display no other health problems until the anaemia observed beyond 14 months of age and otherwise appear normal. The exception is that homozygotes that survive tend to be smaller throughout life, so are unable to make up for this lack of growth in youth.

The survival rate being 52% among homozygous knockout mice implies a range of severity, with variation between mice as to how strong the phenotype is. This may also be why there is a reduction in the genotype incidence but not a complete loss. Behavioural differences are also likely to be involved, with some individual mice feeding more than others do.

The timing of the deaths of the juvenile animals occurring around weaning seems highly suggestive that there is a gastrointestinal cause. Since the drop in survival did not fully track with delayed weaning, shifting by two days rather than seven, this suggests it is not purely due to the change in diet; however, we would expect the pups to naturally increase their consumption of solid food as they age and decrease milk consumption even if they still have access to it. We cannot control for this behaviour and force the pups to only consume milk until 28 days. Furthermore, as this is their natural weaning age, there are many physiological changes that occur in their gastrointestinal system during this period. These include villi and crypt density decreasing after three weeks, with villi and crypt width increasing but few new villi and crypts being formed (Cheng and Bjerknes, 1985); expression of the brushborder enzymes sucrose-isomaltase and lactase-phlorizin hydrolase undergoing significant changes during the suckling-weaning transition around day 17 (Traber, 1999); and multiple claudin regulators of tight junction selectivity showing increased expression from day 14 to day 28 (Holmes et al., 2006).

Neither the low knockout frequency nor the low knockout survival observed in the triple *C11orf* knockout model have been observed in the single *C11orf53* or *C11orf93* knockout lines. Until we establish a *C11orf92* knockout, we cannot know whether they are purely due to the loss of *C11orf92*, or if the effects of each gene loss stack into a more extreme phenotype than is present in any single knockout. If it is the second case, that would imply the genes function in related pathways and provide some degree of redundancy to each other, so creating synthetic lethality. I think it is interesting to note that *C11orf92* is the only gene out of the three that is unique to placental mammals (see figure 1.9), and so if there were to be a single gene with a key role related to gestation and suckling, it would seem a good candidate.

As the aim of this project was to understand not only the function of the genes but also how they were therefore involved in CRC, I was particularly interested in how the knockout mouse model might differ in its development of CRC. However, I was unable to induce a clear CRC phenotype in my model. No growth developed with age, although the animals had to be culled at 14 months old when they began developing anaemia. Due to the highly reduced frequency of *C11orf*$^{-/-}$ mice on the *Apc*$^{Min/+}$ background I have only had one *C11orf*$^{-/-}$ *Apc*$^{Min/+}$ animal, who does appear to have a more extreme phenotype of higher LI polyps and shorter lifespan, but this in only one animal and the LI polyps are still within the observed range of the other genotypes. In both the aging study and use of the susceptible genetic background, while I cannot say there is an increased CRC incidence, other features of the model have prevented my being able to fully explore this.

There is also a survivor bias inherent in the tumour induction studies. The 52% survival at weaning implies a range of severity; as tumour induction takes place in adult mice, it therefore requires selection for those animals with the milder phenotype as the more severely affected animals do not survive to the point at which these studies take place. If a way to improve survival at weaning can be found to allow a higher proportion of the more severely affected mice to reach adulthood, the results in the tumour induction experiments may then change.

Villin is an actin-bundling protein expressed in brushborder epithelial cells within the intestines, and its promoter region can therefore be used to specifically express transgenes within the intestinal epithelium (Pinto et al., 1999). Combining the *Villin* promoter with Cre recombinase allows generation of tissue-specific knockouts of genes flanked by *LoxP* sites

that are restricted to the intestines (El Marjou et al., 2004). This system has been used to generate an alternate *Apc* model, in which somatic mutations of *Apc* are induced in the adult intestines, leaving normal *Apc* function in the rest of the body intact (Hinoi et al., 2007). By crossing the *C11orf* knockout line onto the $Apc^{fl/+}$*/Villin-Cre* background instead of the $Apc^{Min/+}$ background, we can avoid the systemic embryonic interaction between the mutation that so greatly reduces $C11orf^{-/-}$ frequency on the $Apc^{Min/+}$ background.

The *Villin-Cre* system could also be applied to the *C11orf* genes themselves. Restricting the deletion to the intestinal epithelium may reduce the systemic phenotypes seen, such as the reduced frequency at birth and widespread immune system disruption, to give more surviving animals with which to explore an intestinal phenotype. Other possibilities to improve survival include using a heated chamber to raise cage temperature to 25-28°C, and changing the diet to an even higher nutrient content – although, as diet can have a significant effect on intestinal phenotypes, caution would be required if comparing data across different diets.

Alongside the tumour studies, I carried out more general assessment of the *C11orf* knockout line. The blood counts show a reduction in all measured lineages of blood cells, meaning the disruption is at the early stages of haematopoiesis. This corresponds with the reduced spleen size seen on the $Apc^{Min/+}$ background. $Apc^{Min/+}$ mice have altered bone marrow haematopoiesis and so perform extramedullary haematopoiesis, which causes splenomegaly. When there is a *C11orf* mutation, this compensatory mechanism is not as effective, and while their spleens are still enlarged, it is to a lesser degree.

The blood counts give a clue as to the cause of the blood clots seen in the chest cavity. Cervical dislocation can break blood vessels in the neck, hence not being a recommended technique when blood is to be collected. The observed blood in the chest cavity may be due to increased vascular fragility, which could cause more breakage and so allow more blood out of the neck. Alternatively, it may be due to reduced blood clotting, which would give the blood more time to leak into the chest, rather than clotting in the neck at the site of breakage. As the blood counts show a reduction across blood cell lineages, this would support a lack of platelets and so reduction in clotting.

There is strong evidence of immune system disruption in the knockout mice, as alongside the reduction in WBCs there are many immune-related GO terms that are enriched in the differential expression analysis, albeit that this is a preliminary analysis. However, this

corresponds with the link between this locus and CA-CRC. The immune system has a key regulatory role in tissue homeostasis, and disruption of interactions can cause inflammation, which enhances cancer predisposition (De Visser et al., 2006). This interaction is well-established in the large intestine, with IBD being a major risk factor for CRC (Kulaylat and Dayton, 2010; Lutgens et al., 2008).

We have not seen the mice being more susceptible to infection, but they are isolated from many diseases: they are kept in IVCs which filter the air, and although the facility they are housed in is considered "dirty" this is still a relative term, and dirty for a scientific animal unit is still quite clean. A future experiment of interest would be to measure their susceptibility to infection following exposure. If colitis plays a role in these genes' involvement with CRC, infection may also be another method to induce CRC via inflammation, in addition to chemical induction using DSS.

PAS and alcian blue staining shows reduced mucin staining within the intestines of the *C11orf* knockout mice, indicating some dysfunction of the goblet or other secretory cells. Additionally, the Clca1 staining also shows reduction in the mutant mice. The correlates with the RNAscope localisation of mRNA, which indicated that *C11orf53* was being expressed in a specific cell type. These cells were not in the right position within the intestinal crypts to be stem cells or Paneth cells, and so it seemed likely that they were a secretory cell type, as these form the majority of specialised cells in the intestinal epithelium.

The protein staining of reduced mucins and reduced Clca1 contrasts with the expression analysis, which does not show alteration in any key goblet cell markers, including core mucin *Muc2* and *Clca1* itself. One explanation to reconcile these two pieces of data is that tissue staining detects protein, while mRNA-seq detects mRNA. It may be that the goblet cells are still transcribing mucus genes at the same rate but the proteins are not being translated as highly, or are being degraded faster. Goblet cells have extensive ER and quality control systems to ensure the mucus proteins are correctly folded, and any that are not are either retained until they are corrected or are degraded, rather than being secreted. A disruption to protein folding could result in lower protein without reducing mRNA, although markers of ER stress are also unchanged in the expression data. This also fits with the expression data showing reduction across tuft cell markers; goblet cells and tuft cells could be independently affected by the mutation, but as they are both secretory cell types a disruption of this system

would affect both. It is unclear why goblet cells are only affected at the protein level but tuft cells show a reduction at the mRNA level.

The knockout mouse model created here displays a wide-ranging phenotype, and there is still much to be done to identify and understand its various features. I deleted three genes, as all are linked to CRC risk, but that does not mean they all act in the same pathway. The mutation could affect multiple systems and a single explanation for all the data may not be possible. Utilisation of the single gene KO models may help to distinguish between the phenotypes and link with specific gene products.

Chapter 6: Derivation and Utilisation of Ex Vivo Systems from Mouse Tissue

6.1 Introduction

Animal models provide an invaluable research tool, allowing us to assess the effect of an alteration of one element on the entire body and all its systems. However, their complexity can make it challenging to quantify individual features, and there can be practical difficulties – to measure cell proliferation within the intestinal epithelium of a live mouse involves both logistical and ethical issues as to how to carry out such an experiment.

One method for monitoring proliferation within an animal is through use of bromodeoxyuridine (BrDU), which incorporates into DNA in place of thymine (Nowakowski et al., 1989) and so can be used to measure the extent of DNA synthesis between BrDU administration and tissue staining (Drucker et al., 1996). However, BrDU has been shown to have negative side effects, including inducing premature senescence (Michishita et al., 1999) and inhibiting cancer cell proliferation (Levkoff et al., 2008); the extent of these depends on the time period between exposure and tissue collection. Lineage tracing can be carried out using fluorescent proteins under Cre-Lox control to label stem cells and visualise all daughter cells (Schepers et al., 2012; Snippert et al., 2010).

Another alternative is to use animal models as the source of material to establish ex vivo systems. This allows additional means of analysis alongside the whole animal, while preserving a consistent mutation and genetic background across a study. I used mice from the *C11orf*$^{-/-}$ colony to derive primary cell lines, which I then used in proliferation and migration assays. I also worked with Anna Maria Ochocka-Fox to culture intestinal organoids. These three-dimensional tissue cultures replicate the intestinal crypt structure and cell types; their growth is well characterised so we could observe any differences between the genotypes.

6.2 Fibroblast cell lines from mouse ear tissue

6.2.1 Methods

6.2.1.1 Establishment of fibroblast cell cultures

In order to measure the effect of *C11orf* loss on cell activity, I derived primary cell lines from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice, using collagenase-D-pronase digestion of the tissue followed by filtering out individual cells (Khan and Gasser, 2016). An alternative would have been to generate the mutation in an established cell line, such as a human CRC cell line, but we wanted the consistency of identical mutations in the animal, organoid and cell models.

My cell lines were fibroblast cells derived from ear tissue harvested immediately post-mortem, due to their relative ease of derivation and maintenance, but they are neither intestinal nor epithelial and this should be considered when assessing results.

6.2.1.2 Cell activity assays

I used assays to investigate some key features of cell activity that are altered in cancer cells, specifically proliferation and migration.

I measured cell proliferation using an SRB colorimetric assay, in which plates are fixed at set time points and protein stained by SRB dye. The amount of bound dye is quantified by optical plate reader. I plated *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ cells at equal starting density and fixed plates every 24 hours until 96 hours.

I used two kinds of migration assay, both of which involve cells closing a gap in the cell monolayer. Cells were serum-starved to keep cell growth to a minimum and ensure migration was being measured rather than proliferation. I photographed the cells at different time points and used the Measure tool in ImageJ to quantify the remaining area, which was then calculated as a proportion of the original area. Using a script to define the gap was unsuccessful, so the area was defined manually.

The first migration assay I carried out was the scratch assay (figure 6.1a). I cultured *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ cells in 6-well plates to 90% confluency, at which point I switched them to low-serum media. The next day, at 100% confluency, I scratched a line down the centre of

each well using a P200 pipette tip. The cells were washed to remove any loose cells and imaged at 1 hour, 24 hours and 48 hours post scratch.

An issue with the scratch assay is that the action of scratching the cell layer can damage the cells at either edge of the gap, meaning the assay may be measuring the response to damage as well as migration. I therefore also performed a migration assay using chamber inserts, which adhere to the base of 24 well plates (figure 6.1b). Cells are plated in small chambers; when the insert is removed, the cells are in two defined patches with a gap in between, and no damaged cells at the edges. As with the scratch assay, I changed the media to low-serum 24 hours before removing the insert. For this assay we used a live imaging system to photograph the cells every 30 minutes hours for 65 hours, although gap closure analysis was only carried out every five hours.

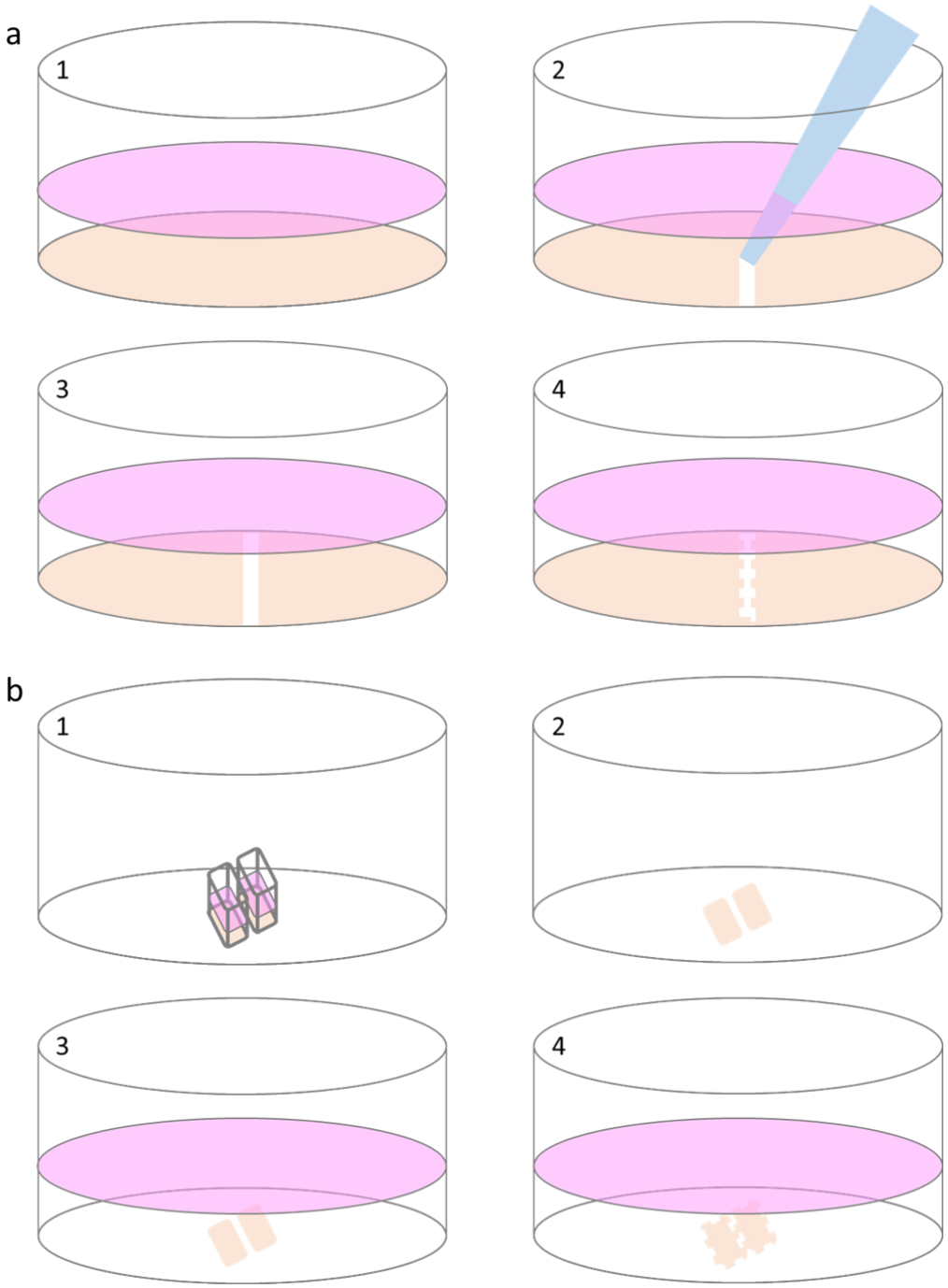


Figure 6.1: (a) Schematic of the scratch migration assay, showing a confluent cell layer being scratched with a pipette tip; cells migrate into the gap left by the scratch and this is photographed and measured. (b) Schematic of the chamber insert migration assay, showing cells growing in adhered chambers which are then removed to leave two distinct patches of cells; cells migrate into the gap between the patches and this is photographed and measured.

6.2.2 Results

I successfully derived primary mouse fibroblast lines from ear tissue harvested from *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ mice (figure 6.2). These cell lines were used in assays for proliferation and migration. No visual differences in morphology were observed between the different genotypes. Cells of different genotypes were extracted and maintained in parallel. I did not immortalise the lines, but only used young cells (passage 4-8) in assays to avoid any effects of aging cells. While I did not use any method of quantification, I observed the cell growth slowed after passage 13.

I carried out the SRB assay twice, each time with 12 replicates per cell line. I also included blank media to subtract background, and an additional outer ring of blank media as an evaporation buffer, as the "edge effect" of 96 well plates can alter cell growth in the out wells. I observed no differences in proliferation rate between *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ cells (figure 6.3).

I performed the scratch assay once (figure 6.4a, 6.4b); each cell line was seeded in two wells, and the scratch in each well was photographed in three places. There is no clear pattern observable in the results (figure 6.5a). At 24 hours there is a significant difference between *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ cells ($p=0.0244$) but this is lost by 48 hours, when all cell lines have almost closed the scratch.

I also ran the chamber insert assay once (figure 6.4c, 6.4d, 6.5b). The cells did not move in a single direction, hence the fluctuations in the proportion of gap remaining as cells retreat backwards. As in the other migration assay, there is no clear indication that one of the genotypes exhibits faster or slower migration.

Unlike in the scratch assay, in the chamber insert assay the gap closure plateaus around 20 hours with a substantial amount of the gap remaining. This may be due to the live imaging we used. Instead of spending the majority of the experiment in an incubator, as the scratched cells did, these cells were in an imaging system for the entire duration. While we did set the temperature and CO_2 to match the incubator (37°C, 5% CO_2), the humidity could not be replicated due to the microscope electrical components and we observed the media level seemed to drop with evaporation. As the plate could not be removed mid-experiment we could not replace this media, and this may have taken a toll on the cells. I attempted to repeat

the assay twice but both times the cells did not show any migration and died before the end of the experiment; this may also have been due to the conditions within the imaging system.

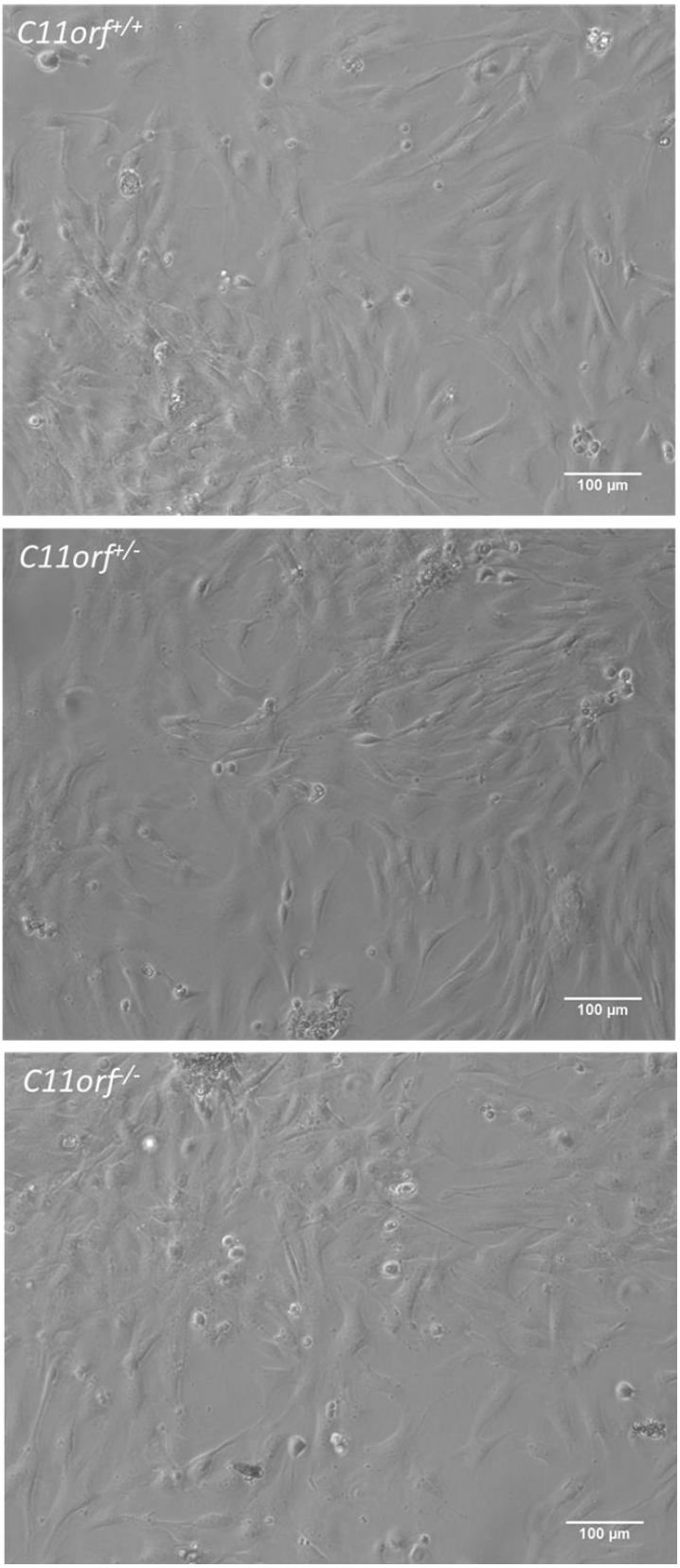


Figure 6.2: Brightfield images of *C11orf*$^{+/+}$, *C11orf*$^{+/-}$ and *C11orf*$^{-/-}$ primary mouse fibroblasts. Fibroblasts were extracted from whole mouse ear tissue and cultured without being immortalised. Photographs taken five days after extraction.

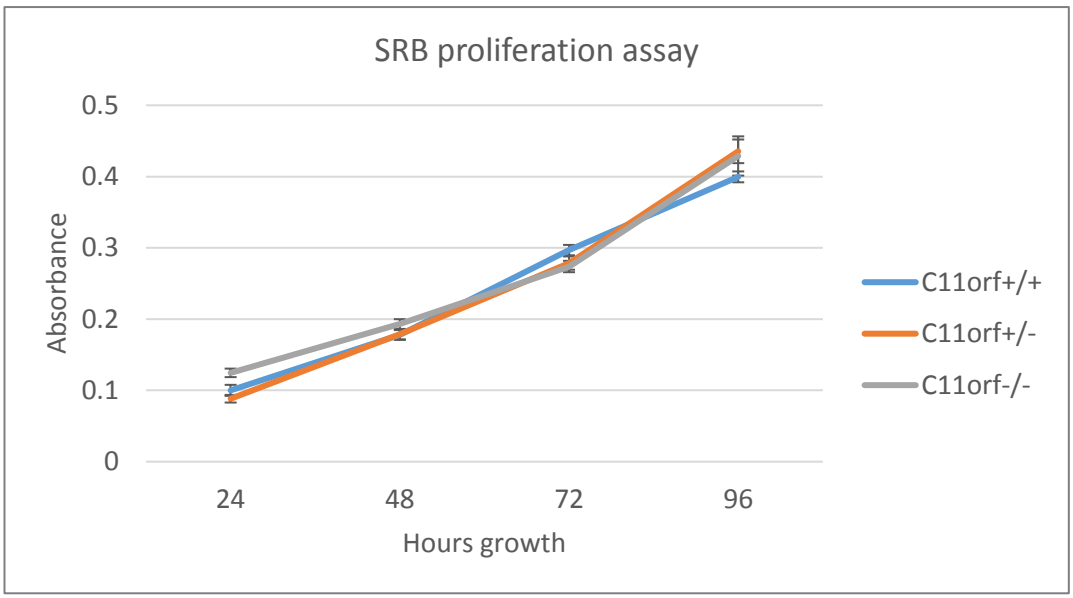


Figure 6.3: SRB proliferation assay on non-immortalised primary mouse fibroblasts at passage 7. Cell lines were plated in 96 well plates at equal density and proliferation was measured every 24 hours by binding of SRB dye, read at 565nm. Experiment contained 24 replicates per cell line.

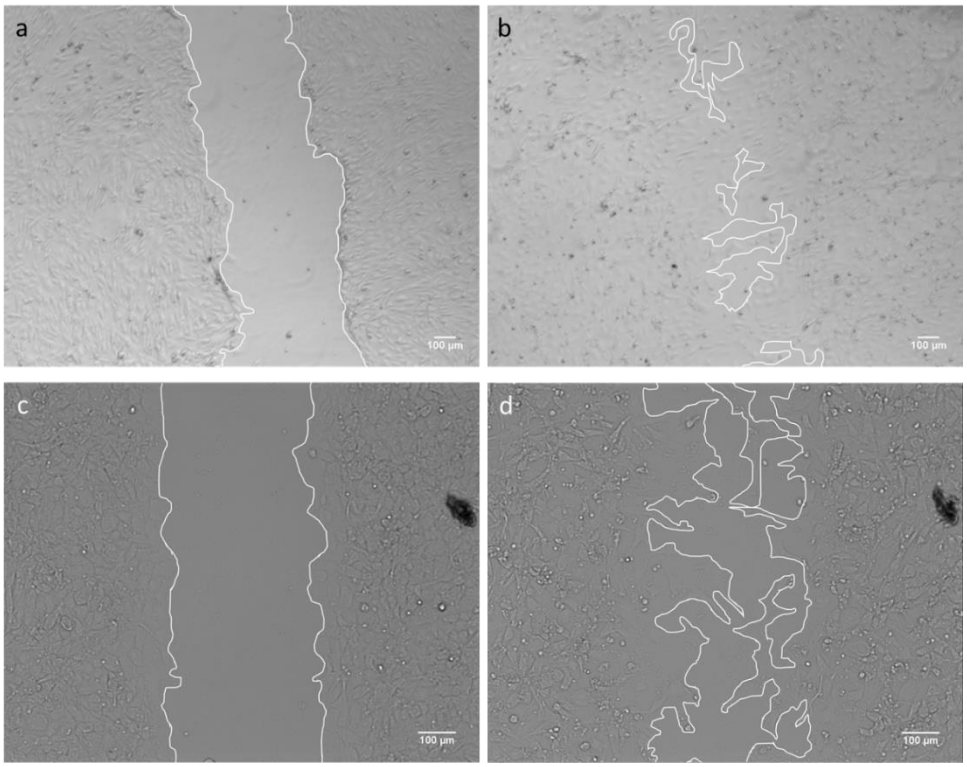


Figure 6.4: Brightfield images of wild type mouse fibroblasts during migration assays: the scratch assay, (a) 1 hour and (b) 48 hours post scratch; the chamber assay, (c) 1 hour and (d) 48 hours post removal of chamber insert. Gap edges have been marked in white to aid visualisation.

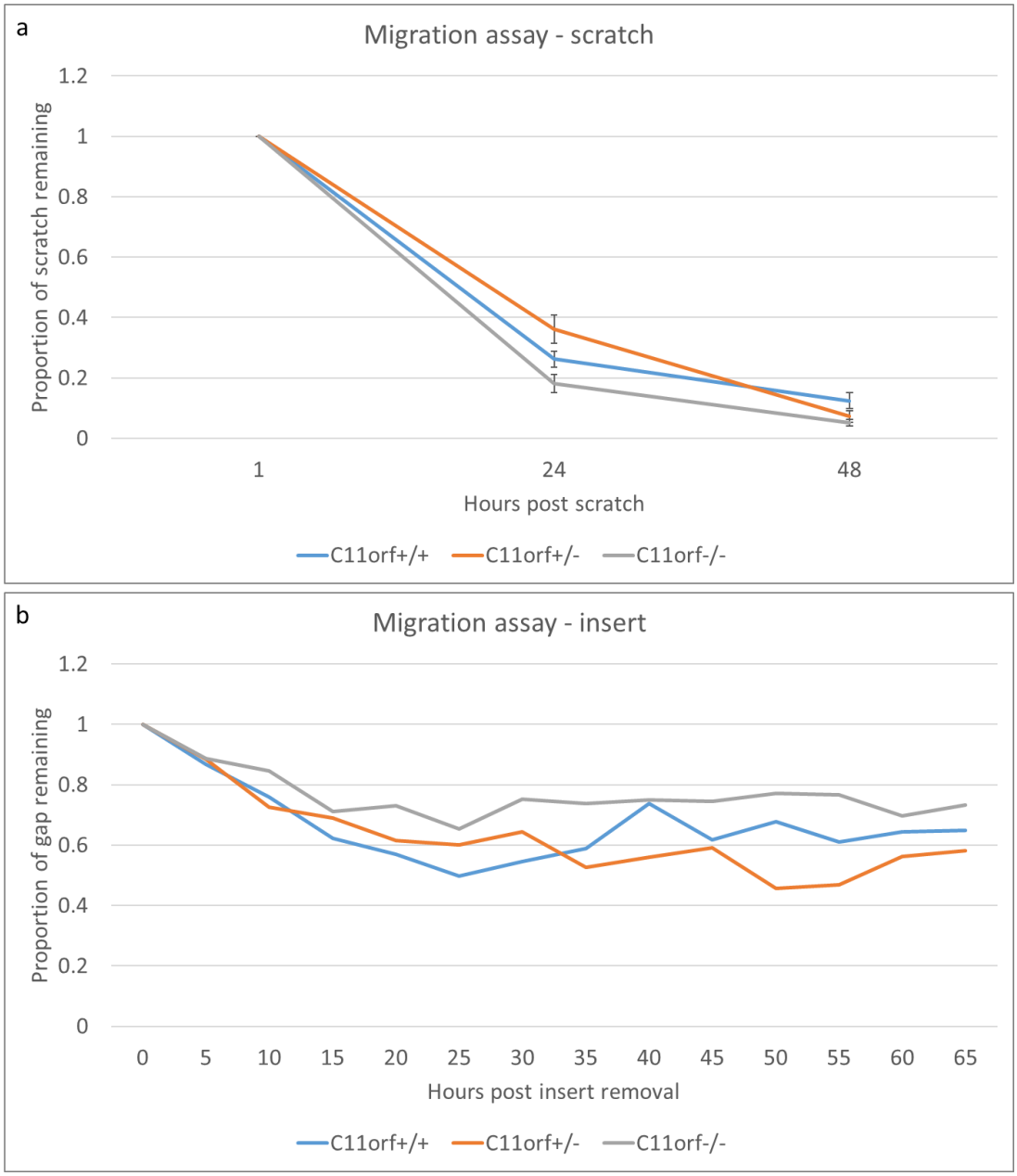


Figure 6.5: Migration assays on primary mouse fibroblasts. (a) The scratch migration assay, performed by scratching a gap into the cell layer, using mouse fibroblasts at passage 4 with six replicates per cell line. (b) The chamber migration assay, performed by culturing cells with an insert to prevent growth within a region, the insert then being removed to leave a gap, using mouse fibroblasts at passage 8 with one replicate per cell line.

6.3 Organoids from mouse intestinal tissue

6.3.1 Methods

Organoids are three-dimensional tissue cultures that are grown in a supportive matrix. Unlike a monolayer of cells in a dish, organoids recapitulate some of the structures found in mature organs, complete with multiple cell types. In the case of intestinal organoids, these "mini guts" consist of complete crypts around a central lumen-like region. Stem cells are isolated from whole tissue and used to generate the organoids; the stem cell niche in the base of the crypts is maintained in mature organoids (Sato et al., 2009). While organoids are more complex than a cell monolayer, they are still an inherently incomplete model, missing the wider context of vasculature and the immune system. This must be remembered when considering results.

Anna Maria Ochocka-Fox derived organoids from freshly harvested proximal small intestine (SI) and large intestine (LI) tissue from *C11orf*$^{+/+}$ and *C11orf*$^{-/-}$ male mice at three weeks and seven weeks old. These were cultured in matrigel with media supplemented with noggin, epidermal growth factor (EGF) and R-spondin, which are required for successful culturing of intestinal organoids (Sato et al., 2009).

6.3.2 Results

Wild type intestinal organoids show a budding phenotype, caused by Paneth cells secreting the growth factor Wnt, which induces proliferation in nearby cells and leads to bud formation (Sato et al., 2011) (figure 6.6a). As organoids are cultured under non-homeostatic conditions, they show exponential growth and continuous budding, which provides a means by which to measure the activity of the organoid population (Farin et al., 2012). When these Wnt gradients are lost, such as through culture with high levels of supplementary Wnt, or if the organoids have a disruption in the Wnt pathway as in *Apc*$^{Min/+}$ tissue, the organoids do not bud but have a round, cystic shape (Sato et al., 2011) (figure 6.6b). The ratio of budding:cystic organoids can therefore be used as a measure to compare two organoid populations.

Eight days after Anna Maria Ochocka-Fox derived organoids from fresh mouse tissue (figure 6.7), I counted the number of budding and the number of cystic organoids per well and took the mean of the percentage budding in each well. The two biological replicates were

combined; although the mice used were at different ages (three weeks and seven weeks) the wild type and homozygous mice were age matched. There was a significant difference in the both the SI and LI, with fewer *C11orf*$^{-/-}$ organoids showing the budding phenotype than the *C11orf*$^{+/+}$ organoids (figure 6.8). This difference – approximately a 25% reduction in both tissues – suggest an altered intestinal homeostasis phenotype and that there is a disruption somewhere in the normal function of the Wnt signalling pathway in the *C11orf*$^{-/-}$ mice.

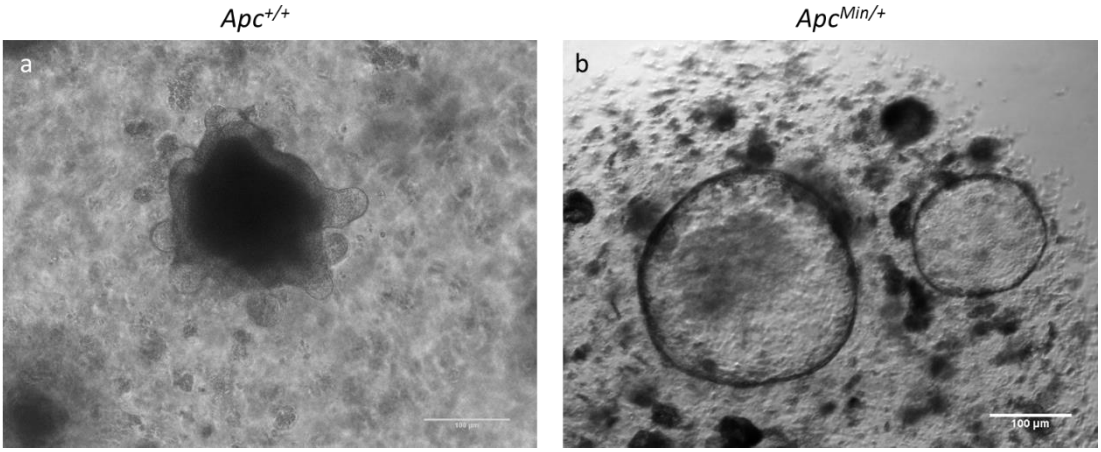


Figure 6.6: Brightfield images of small intestinal organoids grown from *Apc*$^{+/+}$ and *Apc*$^{Min/+}$ mice, photographed on day 8 of culture. (a) Wild type intestinal tissue exhibits the budding phenotype as the organoids recapitulate the crypt structures present in vivo. (b) When the Wnt signalling pathway is disrupted, as in *Apc*$^{Min/+}$ tissue, the budding phenotype is lost to the round cystic phenotype as the lack of Wnt gradients prevents buds from forming. The extent to which an intestinal organoid culture buds can therefore be an indication of how "normal" it is. Images from Anna Maria Ochocka-Fox.

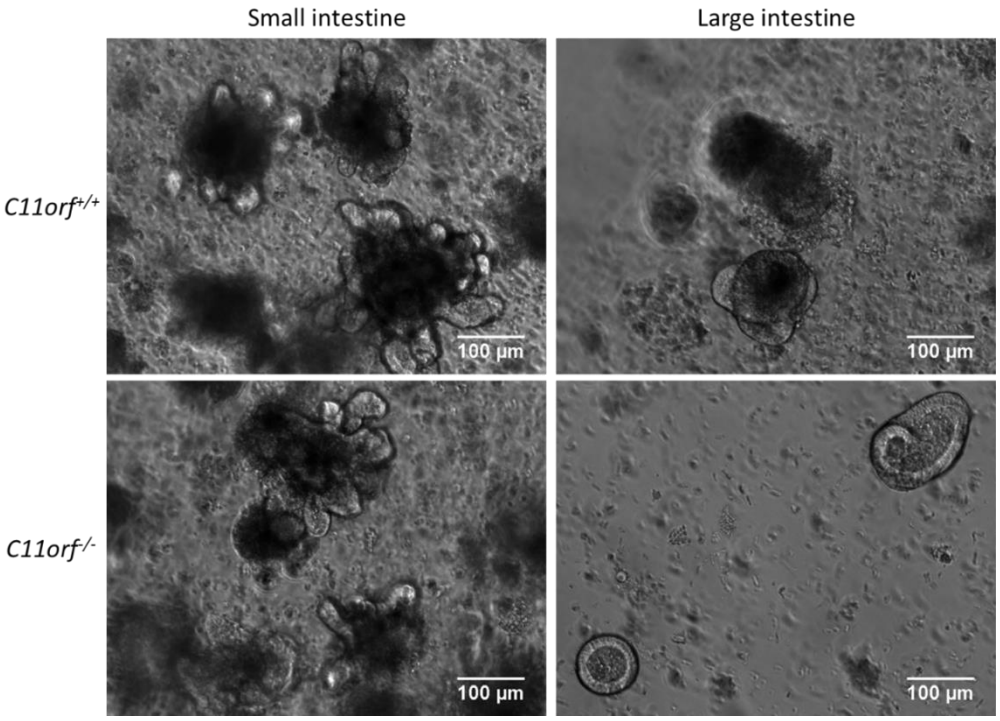


Figure 6.7: Brightfield images of SI and LI organoids grown from *C11orf*$^{+/+}$ and *C11orf*$^{-/-}$ mice, photographed on day 8 of culture. By this day in culture, the budding phenotype is being exhibited by the *C11orf*$^{+/+}$ organoids, and to a lesser extent in the *C11orf*$^{-/-}$ organoids.

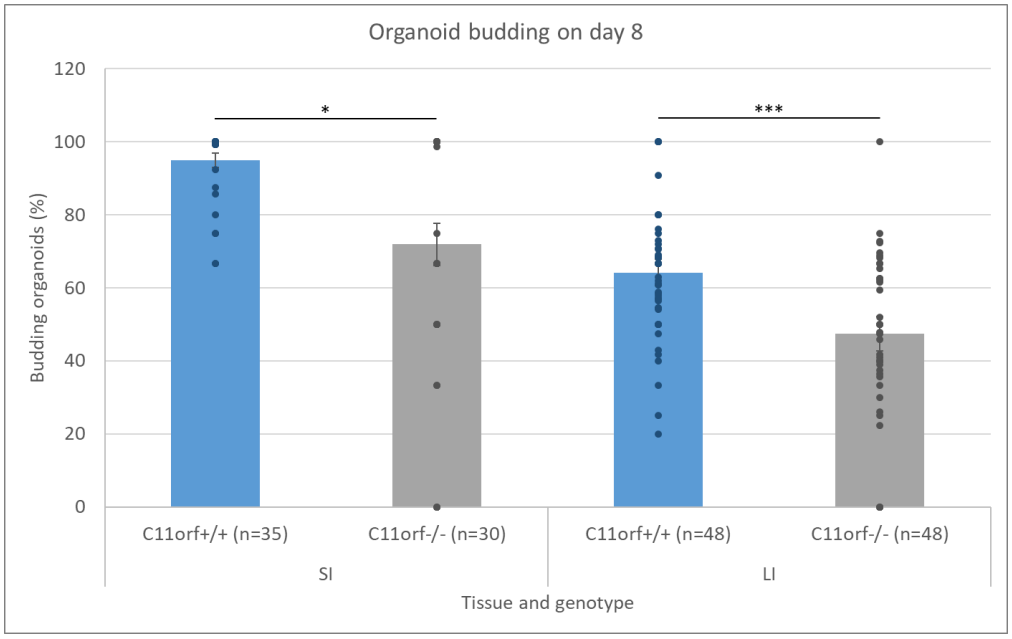


Figure 6.8: Percent of intestinal organoids per well showing the budding phenotype on day 8 of culture. Bars represent the mean, with each well shown as an individual datapoint. Organoids were cultured from SI and LI tissues from *C11orf*$^{+/+}$ and *C1orf*$^{-/-}$ mice. Data was not normally distributed according to Shapiro-Wilk normality test, and so was compared using two-tailed Mann-Whitney U test. Significant results are indicated by stars.

6.4 Discussion

By using the *C11orf*$^{-/-}$ mouse colony as a source of material to set up ex vivo systems, I have been able to quantify aspects of cell activity in a way that would have been highly complex in a living animal. Extracted cell lines have been used to measure cell proliferation and migration, while cultured organoids have given insights into the activity of the Wnt pathway in intestinal tissue. The expression analysis of *C11orf53*, *C11orf92* and *C11orf93* following the GWAS showed that the CRC risk genotype is associated with reduced expression of the three genes, implying they are tumour suppressor genes (Smillie, 2015). I focused on key tumour suppressor mechanisms of proliferation and migration.

The *C11orf*$^{-/-}$ organoids show a significant decrease in budding compared to *C11orf*$^{+/+}$ organoids, both in those derived from the SI and the LI. The size of the effect is far less than that seen in *Apc*$^{Min/+}$ organoids, but the risk conferred by the *C11orf* locus is also far smaller than that of the *Apc* gene.

As the development of buds in organoids has been shown to be due to Wnt gradients, the change in budding observed in the LI organoids indicates some alteration of this Wnt signalling in the *C11orf*$^{-/-}$ mice. This can be through direct activation of the pathway, such as loss of the Wnt antagonist APC (Sato et al., 2011). Aberrant activation of the Wnt pathway has been found in over 90% of human CRC (Cancer Genome Atlas, 2012). This pathway modulates cell proliferation, and it would be expected that an over-activation of the pathway would result in increased cell proliferation.

However, *C11orf*$^{-/-}$ fibroblasts did not show any alteration in proliferation rates compared to *C11orf*$^{+/+}$ and *C11orf*$^{+/-}$ cells. The cells used were fibroblasts extracted from ear tissue, so this does not necessarily mean intestinal epithelial cells do not show a difference, but the fibroblast data do agree with human CRC cell line findings that proliferation is not affected by either *C11orf* siRNA knockdown (Smillie, 2015) or *C11orf* overexpression (see chapter 3). These results together suggest the *C11orf* genes do not function as tumour suppressors through regulation of proliferation rates.

An alternative mechanism behind differences in organoid budding is a change in the degree of differentiation of the crypt cells. Since it is the Paneth cells that secrete Wnt to the surrounding cells, a shift in the population from Paneth cells to stem cells reduces the amount of Wnt that can be produced (Schuijers et al., 2015). The Wnt signalling pathway is also itself

involved in differentiation (Ogaki et al., 2013). As we have not found any evidence of the *C11orf* genes affecting cell proliferation, it may therefore be more likely that the *C11orf*$^{-/-}$ organoids have an increased degree of stemness.

A further inconsistency comes from preliminary expression data from mouse whole intestinal tissue, which does not indicate that Wnt signalling has been affected, with no significant differences in expression of key Wnt pathway components such as *Wnt3*, *Apc*, *Ctnnb1*, *Cdh1* or *Fzd7* between *C11orf* genotypes (see chapter 5 for expression analysis). This does not rule out a disruption in the activity or response to some element of the pathway, beyond their expression levels. We must also consider that pathways may not respond in the same way in fibroblasts, in organoid ex-vivo systems, or in complete animals.

I also investigated cell migration, as another common mechanism for tumour suppressors is inhibition of metastasis, and Wnt signalling has been shown to promote cell migration (Vlad-Fiegen et al., 2012). I did not find differences in the migration activity of *C11orf*$^{-/-}$ fibroblasts, but the conditions of these assays – closing a gap in a monolayer – has limited similarity to the physiological movement of a cancer. A true invasion assay, where cells must penetrate into a collagen matrix, would be a more relevant future experiment.

There are other elements of metastasis that could be investigated, such as angiogenesis. It is worth noting, though, that metastasis affects cancer survival rather than incidence, and while there have been many studies linking the tagging SNP rs3802942 to CRC risk (Hofer et al., 2017; Middeldorp et al., 2009; Tanikawa et al., 2018; Tanskanen et al., 2018; Tenesa et al., 2008; Xiong et al., 2010), rs3802842 has not been found to be associated with CRC survival (Phipps et al., 2012; Smith et al., 2015; Theodoratou et al., 2018).

Tumour suppressor genes also have roles in induction of apoptosis and DNA damage repair. The *C11orf* fibroblast lines could be used in assays for these features, such as the comet assay for DNA damage repair (Collins, 2004) or the Annexin V-Affinity assay for apoptosis (Engeland et al., 1996), to assess if the *C11orf* genes act through these mechanisms.

While there are still questions as to the mechanism, the organoids give quantitative data that there are intestinal differences between *C11orf*$^{+/+}$ and *C11orf*$^{-/-}$ mice. These changes can likely be traced back in some way to the Wnt signalling pathway, which is highly mutated in CRC, although as yet we do not know where in the activation, secretion or response to Wnt the disruption is occurring.

The preliminary expression data indicates several key components of the Wnt signalling pathway are being expressed at the correct levels. Further investigation into the activity of the pathway could include assessing the binding and stabilisation of proteins, such as β-catenin to TCF, through mass spectroscopy or pull down assays (Lowry et al., 2005). The response of the system to upstream elements could be tested by knocking down or overexpressing those proteins and measuring the response, such as in increase in β-catenin phosphorylation in response to Wnt overexpression (Van Noort et al., 2002).

Chapter 7: Discussion

7.1 The function of 11q23.1 and its role in CRC

The aim of this project was to investigate *C11orf53*, *C11orf92* and *C11orf93*, three genes in the 11q23.1 locus, which have been associated with CRC risk. I undertook this using visualisation techniques for both protein and mRNA, creation of a knockout mouse model, and derivation of ex vivo systems from that mouse model.

The markedly reduced survival rate of the *C11orf*$^{-/-}$ mice in the week following weaning, in addition to the disparity between expected and observed ratios of genotypes in this colony, are strong evidence that this locus, while largely uncharacterised, has an important function within the mouse. The timing of the survival drop being in the week following natural weaning would also suggest that the cause may be associated with the gastrointestinal system, concurring with this locus' link to colorectal cancer. Additionally, the distinct expression patterns observed with RNAscope are confirmation that this locus contains three separate genes that, at least to some extent, have individual activity.

It is clear from the data collected from the knockout mouse model that loss of *C11orf53*, *C11orf92* and *C11orf93* has a significant effect on the immune system. This is exhibited in the greatly reduced white blood cell counts measured in the aged *C11orf*$^{-/-}$ mice, and in the genome-wide expression analysis within intestinal tissue. There is widespread alteration in the expression of immune-related genes in the mutant mice, which can be seen in the high number of immunoglobin genes, among other immune-related genes, that occur in the lists of significantly altered genes between genotypes. This is confirmed by gene ontology analysis of these lists, which reports many immune-related gene ontology terms as being significantly enriched. A further link between 11q23.1 and the immune system comes from the locus' position within a wider region that modifies CA-CRC susceptibility (Van Der Kraak et al., 2010).

The immune system has a key role in the initiation and development of cancer, both in prevention and promotion, through mechanisms including modulation of angiogenesis, mediation of tumour cell lysis, and suppression of antitumour immune responses (De Visser et al., 2006). Within the intestines, inflammation is a significant risk factor for CRC, with IBD increasing risk of CRC by up to 70% and being the third highest risk factor for CRC behind the

hereditary conditions FAP and Lynch syndrome (Kulaylat and Dayton, 2010; Lutgens et al., 2013).

An essential part of the immune system within the intestines is the production of mucus, which reduces inflammation and infection by easing passage of digestive material and limiting intestinal epithelium contact with digestive material and microorganisms. Mucus components are secreted by goblet cells and there are many interactions between goblet cells and other elements of the immune system. During infectious colitis, goblet cells produce resistin-like molecule-β which recruits $CD4^+$ T cells (Bergstrom et al., 2015), while deficiency in γδ-T cells results in altered goblet cells and mucin expression (Kober et al., 2014). Tissue staining on mouse intestinal tissue shows that *C11orf*$^{-/-}$ mice have a reduction in production of various mucus components, and so this element of the immune system, in addition to the WBCs and immunoglobins, is also affected by loss of the 11q23.1 locus.

Due to the large volume of proteins being secreted, goblet cells have high ER activity, and disruption of normal ER function disrupts the production of the mucus layer (Zhao et al., 2010). As the mucus prevents inflammation, ER stress and aberrant mucus assembly is therefore associated with inflammation and colitis (Heazlewood et al., 2008; Johansson et al., 2014) and ER stress is a risk factor for IBD (Hosomi et al., 2015; Kaser et al., 2008). ER stress protein XBP1 has also been linked to intestinal tumourigenesis (Niederreiter et al., 2013).

Claire Smillie has previously found evidence that C11ORF53, and to a lesser extent C11ORF93, localises to the ER, in addition to gene ontology analysis on sets of genes with correlating expression to the *C11orf* genes in which ER-associated terms were frequently found to be enriched (Smillie, 2015). I replicated the localisation of C11ORF53 and C11ORF93 to the ER in my own use of overexpression plasmids carrying DDK-tagged proteins. Furthermore, in my RNAscope work *C11orf53* appears to be primarily expressed in a specific cell type, having a tightly clustered distribution rather than the diffuse pattern of *C11orf92* and *C11orf93*. As within the intestinal tissue, many of the specialised cell types are secretory, this may be further evidence of a function within secretory cells. While the immunohistochemical data of reduced mucus staining is contrasted by the expression data, which does not indicate an alteration in key goblet markers including core mucin *Muc2*, the expression data does show reduction in several markers of tuft cells, another secretory cell type.

There is also data that the *C11orf* pathway in some way interacts with the Wnt signalling pathway. The *Apc*$^{Min/+}$ background, which has aberrantly activated Wnt signalling, exhibits both lower expression of the *C11orf* genes and reduced frequency of *C11orf*$^{-/-}$ mice. The reduced budding of the *C11orf*$^{-/-}$ organoids can be linked back to an alteration in Wnt signalling, as it is this pathway which governs the proliferation of intestinal cells and so the budding of intestinal organoids (Sato et al., 2011). Wnt growth factors are secreted by Paneth cells which are found in the base of the intestinal crypts, alongside the stem cells. RNAscope visualisation revealed that *C11orf92* and *C11orf93*, in particular, have their highest expression in the base of the crypts.

The blood analysis shows a reduction across all measured blood cell lineages, with both RBCs and multiple types of WBC having low numbers and so indicating a malfunction at the stem cell level. The Wnt signalling is involved in induction and maintenance of haematopoietic stem cells (Reya et al., 2003; Tran et al., 2010) and *Apc*$^{Min/+}$ mice have disrupted haematopoiesis (Lane et al., 2010; You et al., 2006), and so this phenotype may likewise be attributable to changes in the Wnt signalling pathway.

As they are secreted proteins, members of the Wnt family pass through the ER, where they are retained by proteins such as GRP78 to regulate Wnt signalling (Burrus and McMahon, 1995; Zoltewicz et al., 2009). It has been shown that ER stress and the unfolded protein response alter the stemness of intestinal cells (Heijmans et al., 2013; Niederreiter et al., 2013) and the haematopoietic stem cell pool (Van Galen et al., 2014), both of which are maintained by Wnt signalling (Pinto et al., 2003; Reya et al., 2003).

Given the data from both localisation studies and expression analysis that links the *C11orf* locus to the ER, I propose that disruption of these genes has an effect on the secretory system of cells, which results in reduced mucin production. Wnt signalling is a highly complex pathway and there are many points at which it may interact with *C11orf*. However, the proposed secretory/ER association provides a potential mechanism that underlies both the Wnt and immune data, rather than two separate affected systems, and I therefore suggest that secretion of Wnt has been affected by loss of the *C11orf* genes in the same way as secretion of mucus components.

7.2 Future work

This project has made some progress in characterising the roles of *C11orf53*, *C11orf92* and *C11orf93*, but there are still many unanswered questions. In this section, I will discuss some potential routes for further work, as research in the 11q23.1 locus continues.

Improved knowledge of the localisation of the *C11orf* gene products within the cell and tissue contexts will be key in understanding the genes' mechanism of action. In the continued absence of specific antibodies, there should be further attempts to add endogenous tags to C11ORF53 and C11ORF93, as this will allow both improved visualisation and pull-down or mass spectrometry assays to determine binding partners. Additionally, by using duplex RNAscope or improving the coupling of RNAscope to IHC, *C11orf53* expression should be compared to a panel of markers to determine which cell type is producing the distinctive expression pattern observed, such as *Dclk1* for tuft cells (Gerbe et al., 2012).

The low numbers of knockout mice available has meant much of my data is based on a small number of replicates, or single animals. A priority for further study is therefore to repeat these experiments, such as the differential expression analysis, to generate more robust data that will allow higher confidence in the conclusions drawn. In particular, once there are replicates of the mRNA-seq data, a far higher level of analysis can take place, including considering each region of the intestines individually rather than having to pool all data into simply small intestine and large intestine. This deeper exploration would be valuable as CRC incidence is not consistent throughout the length of the large intestine. In the case of 11q23.1, the risk SNP rs3802842 confers a higher risk of rectal cancer than of colonic cancer (Tenesa et al., 2008).

Obtaining more mice for analysis would be greatly aided by finding a way to improve the survival of the homozygote knockouts, as the alternative is breeding a great many animals with surplus wild types and heterozygotes. This could be done by trying a higher-nutrient diet or increasing the temperature of the cages. Additionally, as the systemic knockout is clearly difficult to cope with at young ages, conditional knockouts in which the mutation is restricted to the intestinal epithelium or induced after one month of age would reduce the severity of the phenotype and so produce more animals which reach maturity.

A reduced severity knockout model may have higher incidence on the $Apc^{Min/+}$ background, allowing better exploration of the effect of *C11orf* loss on a CRC susceptibility background,

including polyp number and lifespan. Alternatively, the existing *C11orf* knockout line could be crossed onto a different susceptible line such as the *Apc*$^{fl/fl}$*/Villin-Cre* background, to reduce the *Apc* phenotype severity rather than the *C11orf* severity.

There is also scope for more work with chemical tumour induction. In addition to the one week, 40mg/kg regime I used, there are multiple other protocols with varying dose and time combinations, and these could be tested on the *C11orf* line. We are also keen to introduce colitis via DSS consumption to the tumour induction studies, in line with the finding that these genes are within a locus which modifies mouse susceptibility to CA-CRC (Van Der Kraak et al., 2010).

The single knockout models of *C11orf53* and *C11orf93* still require much characterisation, *C11orf53* in particular. Thus far, these lines have not displayed the reduced homozygote knockout incidence or survival of the triple gene knockout line; we have yet to test their CRC susceptibility status, or perform many of the other analyses such as blood counts or immunohistochemistry that we have done on the triple knockout line. This data will allow phenotype elements to be assigned to specific genes, if they are only present in one of the single gene knockout lines. Alternatively, if the same phenotype is present in multiple lines, it will imply the genes function at different points of a common pathway. If the triple gene knockout line has features not displayed by any single gene knockout line, this will indicate the genes have an element of redundancy, or function in related pathways and so compound knockouts produce additive effects.

Full exploration of this will also require a *C11orf92* knockout line. Of the 31 *C11orf92* mutations my CRISPR experiment produced, 21 are large deletions between target sites in multiple genes, and so are not restricted to *C11orf92*, with the remaining 10 mutations being small indels. Unlike *C11orf53* and *C11orf93*, which are protein coding and so can be reasonably expected to be disrupted by a frameshift mutation, *C11orf92* is noncoding, and so the effect of a small indel is much harder to predict. These lines would need to be examined via qRT-PCR or similar to confirm a reduction in *C11orf92* expression, and it may be necessary to carry out further CRISPR, perhaps using multiple guides throughout the gene to generate large deletions within *C11orf92*, to produce a *C11orf92* single knockout line with confidence.

With regards to furthering specific results I have obtained, there are methods to investigate the physical properties of mucus following harvesting of mouse intestinal tissue, including

measurement of its thickness, having utilised a charcoal suspension to mark the surface, and measurement of its permeability through tracking the downward progress of fluorescent beads placed on the surface (Gustafsson et al., 2012). These assays can be used to quantify the mucus changes indicated by the tissue staining in the *C11orf* knockout mice.

Additionally, the endoplasmic reticulum function can be investigated in more detail in knockout cell lines – either in my existing mouse fibroblast cells, in a new epithelial line that could be derived from the mouse colony, or in human knockout cell lines that could be generated using CRISPR or another gene editing technique. ER studies could include live imaging using systems such as ER-tracker dyes (Cole et al., 2000), quantification of ER stress using Thioflavin T (Beriault and Werstuck, 2013) or in vitro quantification of ER-associated degradation (Brodsky, 2010).

The blood counts show pancytopenia, with reduction across all blood cell types. As these animals were closely watched and were necropsied, we know that it is not caused by an acute blood loss, and therefore is due to some form of stem cell injury. Blood smears should be carried out for visual assessment of the blood cells, as haematopoietic cell neoplasia of bone marrow or lymphocytic origin presents with abnormal cells in the blood (Darren Wood, 2018). Preliminary examination of bone marrow H&Es have not revealed an obvious change, but they should be further studied, perhaps utilising additional tissue staining or collection of liquid bone marrow rather than fixed whole bones.

Ultimately, while there are many avenues that remain to be explored with regard to these genes, there must be a balance struck between wide-ranging study to cover as much as possible, and a focussed investigation into their role in CRC risk. The future work of this lab will be directed towards the latter, as thus far the 11q23.1 locus has only been linked to CRC risk and so improved understanding of this mechanism offers the most potential for improving public health through advancing CRC prevention, detection and treatment. However, I believe that the wider effects of the 11q23.1 genes have proved to be highly interesting and it is my hope that these will also be further researched in the future.

7.3 Summary

Following a GWAS which identified 11q23.1 as a CRC risk locus, and subsequent expression analysis which implicated largely-uncharacterised genes *C11orf53*, *C11orf92* and *C11orf93*, I have investigated these genes for insights into their function and role in CRC development. Localisation studies have shown that C11ORF53 may localise to the ER and be specifically expressed in a secretory cell type. Use of a knockout mouse has revealed a strong association with the immune system, which is altered in a knockout model with all three genes compromised. While CRC has not been induced, the animals still show an intestinal phenotype, including changes in mucus production within the intestines, in addition to disrupted haematopoiesis and low survival following weaning. Intestinal organoid culture has indicated altered Wnt signalling in the knockout.

I propose that these genes play a role in the secretory system via the ER, and reduction of their activity affects proper secretion components both of intestinal mucus and of the Wnt signalling pathway, in turn producing the observed phenotypes in the mouse model and increasing CRC risk in humans. There is still much that needs to be done to fully understand this locus, but these results give a direction for future work, which will hopefully one day lead to clinical impact.

Bibliography

Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M., and Warusavitarne, J. (2012). Molecular pathways in colorectal cancer. J. Gastroenterol. Hepatol. *27*, 1423–1431.

Alavanja, M., Baron, J.A., Brownson, R.C., Buffler, P.A., DeMarini, D.M., Djordjevic, M. V., Doll, R., Fontham, E.T.H., Gao, Y.T., Gray, N., et al. (2004). Tobacco smoke and involuntary smoking. In IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, pp. 53–119.

Atkin, W.S., and Saunders, B.P. (2002). Surveillance guidelines after removal of colorectal adenomatous polyps. Gut *51*, V6–V9.

Aune, D., Chan, D.S.M., Lau, R., Vieira, R., Greenwood, D.C., Kampman, E., and Norat, T. (2011). Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ *343*, d6617–d6617.

Aune, D., Chan, D.S.M., Vieira, A.R., Navarro Rosenblatt, D.A., Vieira, R., Greenwood, D.C., Kampman, E., and Norat, T. (2013). Red and processed meat intake and risk of colorectal adenomas: a systematic review and meta-analysis of epidemiological studies. Cancer Causes Control *24*, 611–627.

Aust, D.E., Terdiman, J.P., Willenbucher, R.F., Chang, C.G., Molinaro-Clark, A., Baretton, G.B., Loehrs, U., and Waldman, F.M. (2002). The APC/β-catenin pathway in ulcerative colitis-related colorectal carcinomas: A mutational analysis. Cancer *94*, 1421–1427.

Bagnardi, V., Rota, M., Botteri, E., Tramacere, I., Islami, F., Fedirko, V., Scotti, L., Jenab, M., Turati, F., Pasquali, E., et al. (2015). Alcohol consumption and site-specific cancer risk: A comprehensive dose-response meta-analysis. Br. J. Cancer *112*, 580–593.

Bansal, G., DiVietro, J.A., Kuehn, H.S., Rao, S., Nocka, K.H., Gilfillan, A.M., and Druey, K.M. (2008). RGS13 controls G protein-coupled receptor-evoked responses of human mast cells. J. Immunol. *181*, 7882–7890.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003–1007.

Behrens, J., Jerchow, B.-A., Würtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kühl, M., Wedlich, D., and Birchmeier, W. (1998). Functional Interaction of an Axin Homolog, Conductin, with β-Catenin, APC, and GSK3β. Science1 *280*, 596–599.

Ben, Q., Sun, Y., Chai, R., Qian, A., Xu, B., and Yuan, Y. (2014). Dietary fiber intake reduces risk for colorectal adenoma: A meta-analysis. Gastroenterology *146*, 689–699.

Bergeron, J.J.M., Brenner, M.B., Thomas, D.Y., and Williams, D.B. (1994). Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. Trends Biochem. Sci. *19*, 124–128.

Bergstrom, K.S.B., Morampudi, V., Chan, J.M., Bhinder, G., Lau, J., Yang, H., Ma, C., Huang, T., Ryz, N., Sham, H.P., et al. (2015). Goblet Cell Derived RELM-β Recruits CD4+ T Cells during Infectious Colitis to Promote Protective Intestinal Epithelial Cell Proliferation. PLoS Pathog. *11*, e1005108.

Bergström, J.H., Birchenough, G.M.H., Katona, G., Schroeder, B.O., Schütte, A., Ermund, A.,

Johansson, M.E. V., and Hansson, G.C. (2016). Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. Proc. Natl. Acad. Sci. *113*, 13833–13838.

Beriault, D.R., and Werstuck, G.H. (2013). Detection and quantification of endoplasmic reticulum stress in living cells using the fluorescent compound, Thioflavin T. Biochim. Biophys. Acta *1833*, 2293–2301.

Biancolella, M., Fortini, B.K., Tring, S., Plummer, S.J., Mendoza-Fandino, G.A., Hartiala, J., Hitchler, M.J., Yan, C., Schumacher, F.R., Conti, D. V, et al. (2014). Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23.1. Hum. Mol. Genet. *23*, 2198–2209.

Birchenough, G.M.H., Johansson, M.E., Gustafsson, J.K., Bergström, J.H., and Hansson, G.C. (2015). New developments in goblet cell mucus secretion and function. Mucosal Immunol. *8*, 712–719.

Bird, R.P. (1987). Observation and quantification of aberrant crypts in the murine colon treated with a colon carcinogen: Preliminary findings. Cancer Lett. *37*, 147–151.

Boland, R.C., and Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. Gastroenterology *138*, 2073–2087.

Bowcutt, R., Forman, R., Glymenaki, M., Carding, S.R., Else, K.J., and Cruickshank, S.M. (2014). Heterogeneity across the murine small and large intestine. World J. Gastroenterol. *20*, 15216–15232.

Boyle, T., Keegel, T., Bull, F., Heyworth, J., and Fritschi, L. (2012). Physical activity and risks of proximal and distal colon cancers: A systematic review and meta-analysis. J. Natl. Cancer Inst. *104*, 1548–1561.

Brabletz, T., Hlubek, F., Spaderna, S., Schmalhofer, O., Hiendlmeyer, E., Jung, A., and Kirchner, T. (2005). Invasion and Metastasis in Colorectal Cancer: Epithelial-Mesenchymal Transition, Mesenchymal-Epithelial Transition, Stem Cells and β-Catenin. Cells Tissues Organs *179*, 56–65.

Brannon, A.R., Vakiani, E., Sylvester, B.E., Scott, S.N., McDermott, G., Shah, R.H., Kania, K., Viale, A., Oschwald, D.M., Vacic, V., et al. (2014). Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. Genome Biol. *15*, 454.

Brodsky, J.L. (2010). The use of in vitro assays to measure endoplasmic reticulum-associated degradation. Methods Enzymol. *470*, 661–679.

ten Broeke, S.W., Elsayed, F.A., Pagan, L., Olderode-Berends, M.J.W., Garcia, E.G., Gille, H.J.P., van Hest, L.P., Letteboer, T.G.W., van der Kolk, L.E., Mensenkamp, A.R., et al. (2017). SNP association study in PMS2-associated Lynch syndrome. Fam. Cancer *0*, 1–9.

Brown, K.F., Rumgay, H., Dunlop, C., Ryan, M., Quartly, F., Cox, A., Deas, A., Elliss-Brookes, L., Gavin, A., Hounsome, L., et al. (2018). The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br. J. Cancer *118*, 1130–1141.

Burrus, L.W., and McMahon, A.P. (1995). Biochemical Analysis of Murine Wnt Proteins Reveals both Shared and Distinct Properties. Exp. Cell Res. *220*, 363–373.

Bustin, S.A., Li, S.-R., and Dorudi, S. (2001). Expression of the Ca2+-activated chloride channel genes CLCA1 and CLCA2 is downregulated in human colorectal cancer. DNA Cell Biol. *20*, 331–338.

Canavan, C., Abrams, K.R., and Mayberry, J. (2006). Meta-analysis: Colorectal and small bowel cancer risk in patients with Crohn's disease. Aliment. Pharmacol. Ther. *23*, 1097–1104.

Cancer Genome Atlas (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Cancer Research UK (2016). Bowel cancer statistics, https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer. Accessed July 2018.

Carvajal-Carmona, L.G., Zauber, A.G., Jones, A.M., Howarth, K., Wang, J., Cheng, T., Riddell, R., Lanas, A., Morton, D., Bertagnolli, M.M., et al. (2013). Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. Gastroenterology *144*, 53–55.

Castaño-Milla, C., Chaparro, M., and Gisbert, J.P. (2014). Systematic review with meta-Analysis: The declining risk of colorectal cancer in ulcerative colitis. Aliment. Pharmacol. Ther. *39*, 645–659.

Chan, D.S.M., Lau, R., Aune, D., Vieira, R., Greenwood, D.C., Kampman, E., and Norat, T. (2011). Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. PLoS One *6*, e20456.

Chang, C.L., Marra, G., Chauhan, D.P., Ha, H.T., Chang, D.K., Ricciardiello, L., Randolph, A., Carethers, J.M., and Boland, C.R. (2002). Oxidative stress inactivates the human DNA mismatch repair system. Am. J. Physiol. Physiol. *283*, C148–C154.

Chen, Q., Wang, J., Yang, J., Jin, Z., Shi, W., Qin, Y., Yu, F., and He, J. (2015). Association between adult weight gain and colorectal cancer: A dose-response meta-analysis of observational studies. Int. J. Cancer *136*, 2880–2889.

Chen, X., Wang, Y., Xia, H., Wang, Q., Jiang, X., Lin, Z., Ma, Y., Yang, Y., and Hu, M. (2012). Loss of E-cadherin promotes the growth, invasion and drug resistance of colorectal cancer cells and is associated with liver metastasis. Mol. Biol. Rep. *39*, 6707–6714.

Chen, X., Zaro, J.L., and Shen, W.C. (2013a). Fusion protein linkers: Property, design and functionality. Adv. Drug Deliv. Rev. *65*, 1357–1369.

Chen, Z., He, X., Jia, M., Liu, Y., Qu, D., Wu, D., Wu, P., Ni, C., Zhang, Z., Ye, J., et al. (2013b). β-catenin Overexpression in the Nucleus Predicts Progress Disease and Unfavourable Survival in Colorectal Cancer: A Meta-Analysis. PLoS One *8*, e63854.

Cheng, H., and Bjerknes, M. (1985). Whole Population Cell Kinetics and Postnatal Development of the Mouse Intestinal Epithelium. Anat. Rec. *211*, 420–426.

Cheng, K., Samimi, R., Xie, G., Shant, J., Drachenberg, C., Wade, M., Davis, R.J., Nomikos, G., Raufman, J.-P., Rj, D., et al. (2008). Acetylcholine release by human colon cancer cells mediates autocrine stimulation of cell proliferation. Am J Physiol Gastrointest Liver Physiol *295*, G591–G597.

Cheng, T.H., Gorman, M., Martin, L., Barclay, E., Casey, G., Colon Cancer Family Registry, CGEMS, Saunders, B., Thomas, H., Clark, S., et al. (2015). Common colorectal cancer risk alleles contribute to the multiple colorectal adenoma phenotype, but do not influence colonic polyposis in FAP. Eur. J. Hum. Genet. *23*, 260–263.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. *24*, 132–141.

Closa, A., Cordero, D., Sanz-Pamplona, R., Solé, X., Crous-Bou, M., Paré-Brunet, L., Berenguer, A., Guino, E., Lopez-Doriga, A., Guardiola, J., et al. (2014). Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. Carcinogenesis *35*, 2039–2046.

Cole, L., Davies, D., Hyde, G.J., and Ashford, A.E. (2000). ER-tracker dye and BODIPY-brefeldin A differentiate the endoplasmic reticulum and Golgi bodies from the tubular-vacuole system in living hyphae of Pisolithus tinctorius. J. Microsc. *197*, 239–248.

Collins, A.R. (2004). The comet assay for DNA damage and repair: principles, applications, and limitations. Mol. Biotechnol. *26*, 249–261.

Cong, F., Schweizer, L., Chamorro, M., and Varmus, H. (2003). Requirement for a Nuclear Function of beta-Catenin in Wnt Signaling. Mol. Cell. Biol. *23*, 8462–8470.

Cong, Y.J., Gan, Y., Sun, H.L., Deng, J., Cao, S.Y., Xu, X., and Lu, Z.X. (2014). Association of sedentary behaviour with colon and rectal cancer: A meta-analysis of observational studies. Br. J. Cancer *110*, 817–826.

Cunningham, D., Atkin, W., Lenz, H.J., Lynch, H.T., Minsky, B., Nordlinger, B., and Starling, N. (2010). Colorectal cancer. Lancet *375*, 1030–1047.

D'amico, M.A., Ghinassi, B., Izzicupo, P., Manzoli, L., and Di Baldassarre, A. (2014). Biological function and clinical relevance of chromogranin A and derived peptides. Endocr. Connect. *3*, R45-54.

Darren Wood, R. (2018). Leukogram Abnormalities https://www.msdvetmanual.com/circulatory-system/leukocyte-disorders/leukogram-abnormalities. Accessed August 2018.

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat. Rev. Cancer *17*, 79–92.

Drake, T.M., Ritchie, J.E., Kanthou, C., Staves, J.J., Narramore, R., and Wyld, L. (2015). Targeting the endoplasmic reticulum mediates radiation sensitivity in colorectal cancer. Exp. Mol. Pathol. *98*, 532–539.

Drees, F., Pokutta, S., Yamada, S., Nelson, W.J., and Weis, W.I. (2005). α-catenin is a molecular switch that binds E-cadherin-β-catenin and regulates actin-filament assembly. Cell *123*, 903–915.

Drucker, D.J., Erlich, P., Asa, S.L., and Brubaker, P.L. (1996). Induction of intestinal epithelial proliferation by glucagon-like peptide 2. Proc. Natl. Acad. Sci. *93*, 7911–7916.

Dunlop, M., Dobbins, S., Farrington, S., Jones, A., Palles, C., Whiffin, N., Tenesa, A., Spain, S.,

Broderick, P., Ooi, L., et al. (2016). Common variation near CDKN1A, POLD3, and SHROOM2 influences colorectal cancer risk. Nat. Genet. *44*, 770–776.

Dunlop, M.G., Tenesa, A., Farrington, S.M., Ballereau, S., Brewster, D.H., Koessler, T., Pharoah, P., Schafmayer, C., Hampe, J., Völzke, H., et al. (2012). Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals. Gut *62*, 871–881.

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. PLoS Comput. Biol. *3*, e39.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 1471–2105.

Elzagheid, A., Emaetig, F., Alkikhia, L., Buhmeida, A., Syrjanen, K., El-Faitori, O., Latto, M., Collan, Y., and Pyrhonen, S. (2013). High cyclooxygenase-2 expression is associated with advanced stages in colorectal cancer. Anticancer Res. *33*, 3137–3144.

Engeland, M. Van, Ramaekers, F.C.S., van Engeland, M., Nieland, L.J.W., Ramaekers, F.C.S., Schutte, B., Reutelingsperger, C.P.M., Casciola-Rosen, L., Rosen, A., Petri, M., et al. (1996). A review on an apoptosis detection system based on phosphatidyserine exposure. Proc. Natl. Acad. Sci. *93*, 1–9.

Fang, J.Y., and Richardson, B.C. (2005). The MAPK signalling pathways and colorectal cancer. Lancet Oncol *6*, 322–327.

Farin, H.F., Van Es, J.H., and Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of paneth cells. Gastroenterology *143*, 1518–1529.

Farrington, S.M., Tenesa, A., Barnetson, R., Wiltshire, A., Prendergast, J., Porteous, M., Campbell, H., and Dunlop, M.G. (2005). Germline susceptibility to colorectal cancer due to base-excision repair gene defects. Am. J. Hum. Genet. *77*, 112–119.

Fedirko, V., Tramacere, I., Bagnardi, V., Rota, M., Scotti, L., Islami, F., Negri, E., Straif, K., Romieu, I., La Vecchia, C., et al. (2011). Alcohol drinking and colorectal cancer risk: An overall and dose-Response meta-analysis of published studies. Ann. Oncol. *22*, 1958–1972.

Fiala, E.S. (1977). Investigations into the metabolism and mode of action of the colon carcinogens 1,2-dimethylhydrazine and azoxymethane. Cancer *40*, 2436–2445.

Friedenreich, C.M. (2011). Physical activity and breast cancer: Review of the epidemiologic evidence and biologic mechanisms. Recent Results Cancer Res. *188*, 125–139.

Van Galen, P., Kreso, A., Mbong, N., Kent, D.G., Fitzmaurice, T., Chambers, J.E., Xie, S., Laurenti, E., Hermans, K., Eppert, K., et al. (2014). The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. Nature *510*, 268–272.

Galiatsatos, P., and Foulkes, W.D. (2006). Familial adenomatous polyposis. Am. J. Gastroenterol. *101*, 385–398.

Gerbe, F., Legraverend, C., and Jay, P. (2012). The intestinal epithelium tuft cells: Specification and function. Cell. Mol. Life Sci. *69*, 2907–2917.

Gerdes, J., Lemke, H., Baisch, H., Wacker, H., Schwab, U., and Stein, H. (1984). Cell cycle

analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. J. Immunol. *133*, 1710–1715.

Glebov, O.K., Rodriguez, L.M., Nakahara, K., Expression, G., Jenkins, J., Cliatt, J., Humbyrd, C., Denobile, J., Soballe, P., Simon, R., et al. (2003). Distinguishing Right from Left Colon by the Pattern of Gene Expression. Cancer Epidemiol. Biomarkers Prev. *12*, 755–762.

González-Sancho, J.M., Aguilera, O., García, J.M., Pendás-Franco, N., Peña, C., Cal, S., De Herreros, A.G., Bonilla, F., and Muñoz, A. (2005). The Wnt antagonist DICKKOPF-1 gene is a downstream target of β-catenin/TCF and is downregulated in human colon cancer. Oncogene *24*, 1098–1103.

Graham, J., Mushin, M., and Kirkpatrick, P. (2004). Oxaliplatin. Nat. Rev. Drug Discov. *3*, 11–12.

Gregorieff, A., and Clevers, H. (2005). Wnt signaling in the intestinal epithelium : from endoderm to cancer Wnt signaling in the intestinal epithelium : from endoderm to cancer. Gene Expr. Patterns *19*, 877–890.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and Van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356.

Gustafsson, J.K., Ermund, A., Johansson, M.E. V., Schutte, A., Hansson, G.C., and Sjovall, H. (2012). An ex vivo method for studying mucus formation, properties, and thickness in human colonic biopsies and mouse small and large intestinal explants. AJP Gastrointest. Liver Physiol. *302*, G430–G438.

Häfner, M.F., and Debus, J. (2016). Radiotherapy for colorectal cancer: Current standards and future perspectives. Visc. Med. *32*, 172–177.

Hamilton, S.R., Vogelstein, B., Kudo, S., Riboli, E., Nakamura, S., Hainaut, P., Rubio, C.A., Sobin, L.H., Fogt, F., Winawer, S.J., et al. (2000). Carcinoma of the colon and rectum. In World Health Organisation Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System, S.R. Hamilton, and L.A. Aaltonen, eds. (Lyon: IARC Press), pp. 103–119.

Hasnain, S.Z., Tauro, S., Das, I., Tong, H., Chen, A.H., Jeffery, P.L., McDonald, V., Florin, T.H., and McGuckin, M.A. (2013). IL-10 promotes production of intestinal mucus by suppressing protein misfolding and endoplasmic reticulum stress in goblet cells. Gastroenterology *144*, 357–368.

Heazlewood, C.K., Cook, M.C., Eri, R., Price, G.R., Tauro, S.B., Taupin, D., Thornton, D.J., Chin, W.P., Crockford, T.L., Cornall, R.J., et al. (2008). Aberrant mucin assembly in mice causes endoplasmic reticulum stress and spontaneous inflammation resembling ulcerative colitis. PLoS Med. *5*, e54.

Heijmans, J., van Lidth de Jeude, J.F., Koo, B.-K., Rosekrans, S.L., Wielenga, M.C.B., van de Wetering, M., Ferrante, M., Lee, A.S., Onderwater, J.J.M., Paton, J.C., et al. (2013). ER Stress Causes Rapid Loss of Intestinal Epithelial Stemness through Activation of the Unfolded

Protein Response. Cell Rep. *3*, 1128–1139.

Hino, K., Saito, A., Asada, R., Kanemoto, S., and Imaizumi, K. (2014). Increased susceptibility to dextran sulfate sodium-induced colitis in the endoplasmic reticulum stress transducer OASIS deficient mice. PLoS One *9*, 1–2.

Hinoi, T., Akyol, A., Theisen, B.K., Ferguson, D.O., Greenson, J.K., Williams, B.O., Cho, K.R., and Fearon, E.R. (2007). Mouse model of colonic adenoma-carcinoma progression based on somatic Apc inactivation. Cancer Res. *67*, 9721–9730.

Hofer, P., Hagmann, M., Brezina, S., Dolejsi, E., Mach, K., Leeb, G., Baierl, A., Buch, S., Sutterlüty-Fall, H., Karner-Hanusch, J., et al. (2017). Bayesian and frequentist analysis of an Austrian genome-wide association study of colorectal cancer and advanced adenomas. Oncotarget *8*, 98623–98634.

Holmes, J.L., Van Itallie, C.M., Rasmussen, J.E., and Anderson, J.M. (2006). Claudin profiling in the mouse during postnatal intestinal development and along the gastrointestinal tract reveals complex expression patterns. Gene Expr. Patterns *6*, 581–588.

Hosomi, S., Kaser, A., and Blumberg, R.S. (2015). Role of endoplasmic reticulum stress and autophagy as interlinking pathways in the pathogenesis of inflammatory bowel disease. Curr. Opin. Gastroenterol. *31*, 81–88.

Houlle, S., Charbonnier, F., Houivet, E., Tinat, J., Buisine, M.P., Caron, O., Benichou, J., Baert-Desurmont, S., and Frebourg, T. (2011). Evaluation of Lynch syndrome modifier genes in 748 MMR mutation carriers. Eur. J. Hum. Genet. *19*, 887–892.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hülsken, J., Birchmeier, W., and Behrens, J. (1994). E-cadherin and APC compete for the interaction with β-catenin and the cytoskeleton. J. Cell Biol. *127*, 2061–2069.

Hussain, S.P., Amstad, P., Raja, K., Hussain, S.P., Amstad, P., Raja, K., Ambs, S., Nagashima, M., Bennett, W.P., Shields, P.G., et al. (2000). Increased p53 Mutation Load in Noncancerous Colon Tissue from Ulcerative Colitis : A Cancer-prone Chronic Inflammatory Disease Advances in Brief Increased p53 Mutation Load in Noncancerous Colon Tissue from Ulcerative Colitis : A Cancer-prone Chronic Infl. Cancer Res. *60*, 3333–3337.

Huxley, R.R., Ansary-Moghaddam, A., Clifton, P., Czernichow, S., Parr, C.L., and Woodward, M. (2009). The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence. Int. J. Cancer *125*, 171–180.

Jacoby, R.F., Llor, X., Teng, B.B., Davidson, N.O., and Brasitus, T.A. (1991). Mutations in the K-ras oncogene induced by 1,2-dimethylhydrazine in preneoplastic and neoplastic rat colonic mucosa. J. Clin. Invest. *87*, 624–630.

Jass, J.R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology *50*, 113–130.

Jiang, Y., Ben, Q., Shen, H., Lu, W., Zhang, Y., and Zhu, J. (2011). Diabetes mellitus and incidence and mortality of colorectal cancer: a systematic review and meta-analysis of cohort studies. Eur J Epidemiol *26*, 863–876.

Johansson, M.E. V, Thomsson, K.A., and Hansson, G.C. (2009). Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the fcgbp protein. J. Proteome Res. *8*, 3549–3557.

Johansson, M.E. V, Gustafsson, J.K., Holmen-Larsson, J., Jabbar, K.S., Xia, L., Xu, H., Ghishan, F.K., Carvalho, F.A., Gewirtz, A.T., Sjovall, H., et al. (2014). Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. Gut *63*, 281–291.

Joosen, A.M.C.P., Kuhnle, G.G.C., Aspinall, S.M., Barrow, T.M., Lecommandeur, E., Azqueta, A., Collins, A.R., and Bingham, S.A. (2009). Effect of processed and red meat on endogenous nitrosation and DNA damage. Carcinogenesis *30*, 1402–1407.

Kaemmerer, E., Plum, P., Klaus, C., Weiskirchen, R., Liedtke, C., Adolf, M., Schippers, A., Wagner, N., Reinartz, A., and Gassler, N. (2010). Fatty acid binding receptors in intestinal physiology and pathophysiology. World J. Gastrointest. Pathophysiol. *1*, 147–153.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature *502*, 333–339.

Kaser, A., Lee, A., Franke, A., and Glickman, J. (2008). XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell *134*, 743–756.

Kastenhuber, E.R., and Lowe, S.W. (2017). Putting p53 in Context. Cell *170*, 1062–1078.

Khan, M., and Gasser, S. (2016). Generating Primary Fibroblast Cultures from Mouse Ear and Tail Tissues. J. Vis. Exp. e53565.

Kim, H.S., Kang, S.H., Park, C.H., Yang, W.I., Jeung, H.C., Chung, H.C., Roh, J.K., Ahn, J.B., Kim, N.K., Min, B.S., et al. (2011). Genome-wide molecular characterization of mucinous colorectal adenocarcinoma using cDNA microarray analysis. Oncol. Rep. *25*, 717–727.

Kimura, H., Hokari, R., Miura, S., Shigematsu, T., Hirokawa, M., Akiba, Y., Kurose, I., Higuchi, H., Fujimori, H., Tsuzuki, Y., et al. (1998). Incraesed expression of an inducible isoform of nitric oxide synthase and the formation of peroxynitrite in colonic mucosa of patients with active ulcerative colitis. Gut. *42*, 180–187.

Klaunig, J.E., Xu, Y., Isenberg, J.S., Bachowski, S., Kolaja, K.L., Jiang, J., Stevenson, D.E., and Walborg, E.F. (1998). The role of oxidative stress in chemical carcinogenesis. Environ. Health Perspect. *106*, 289–295.

Kobayashi, M., Honma, T., Matsuda, Y., Suzuki, Y., Narisawa, R., Ajioka, Y., and Asakura, H. (2000). Nuclear translocation of beta-catenin in colorectal cancer. Br. J. Cancer *82*, 1689–1693.

Kober, O.I., Ahl, D., Pin, C., Holm, L., Carding, S.R., and Juge, N. (2014). γδ T-cell-deficient mice show alterations in mucin expression, glycosylation, and goblet cells but maintain an intact mucus layer. Am J Physiol Gastrointest Liver Physiol *306*, G582–G593.

Van Der Kraak, L., Meunier, C., Turbide, C., Jothy, S., Gaboury, L., Marcus, V., Chang, S.Y., Beauchemin, N., and Gros, P. (2010). A Two-Locus System Controls Susceptibility to Colitis-Associated Colon Cancer in Mice. Oncotarget *1*, 436–446.

Van Kruijsdijk, R.C.M., Van Der Wall, E., and Visseren, F.L.J. (2009). Obesity and cancer: The role of dysfunctional adipose tissue. Cancer Epidemiol. Biomarkers Prev. *18*.

Kulaylat, M.N., and Dayton, M.T. (2010). Ulcerative colitis and cancer. J. Surg. Oncol. *101*, 706–712.

Kupfer, S.S., Anderson, J.R., Hooker, S., Skol, A., Kittles, R.A., Keku, T.O., Sandler, R.S., and Ellis, N.A. (2010). Genetic heterogeneity in colorectal cancer associations between African and European Americans. Gastroenterology *139*, 1677–1685.

Lane, S.W., Sykes, S.M., Al-Shahrour, F., Shterental, S., Paktinat, M., Lo Celso, C., Jesneck, J.L., Ebert, B.L., Williams, D.A., and Gilliland, D.G. (2010). The Apcmin mouse has altered hematopoietic stem cell function and provides a model for MPD/MDS. Blood *115*, 3489–3497.

Larsson, S.C., Orsini, N., and Wolk, A. (2005). Diabetes mellitus and risk of colorectal cancer: A meta-analysis. J. Natl. Cancer Inst. *97*, 1679–1687.

Levkoff, L.H., Marshall, G.P., Ross, H.H., Caldeira, M., Reynolds, B.A., Cakiroglu, M., Mariani, C.L., Streit, W.J., and Laywell, E.D. (2008). Bromodeoxyuridine Inhibits Cancer Cell Proliferation In Vitro and In Vivo. Neoplasia *10*, 804–816.

Li, K., Wang, G., Andersen, T., Zhou, P., and Pu, W.T. (2014). Optimization of genome engineering approaches with the CRISPR/Cas9 system. PLoS One *9*, e105779.

Liang, P.S., Chen, T.-Y., and Giovannucci, E. (2009). Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis. Int. J. Cancer *124*, 2406–2415.

Lim, S.C., Cho, H., Lee, T.B., Choi, C.H., Min, Y.D., Kim, S.S., and Kim, K.J. (2010). Impacts of cytosolic phospholipase A2, 15-prostaglandin dehydrogenase, and cyclooxygenase-2 expressions on tumor progression in colorectal cancer. Yonsei Med. J. *51*, 692–699.

Linnekamp, J.F., Van Hooff, S.R., Prasetyanti, P.R., Kandimalla, R., Buikhuisen, J.Y., Fessler, E., Ramesh, P., Lee, K.A.S.T., Bochove, G.G.W., De Jong, J.H., et al. (2018). Consensus molecular subtypes of colorectal cancer are recapitulated in in vitro and in vivo models. Cell Death Differ. *25*, 616–633.

Lipka, J., Kapitein, L.C., Jaworski, J., and Hoogenraad, C.C. (2016). Microtubule-binding protein doublecortin-like kinase 1 (DCLK1) guides kinesin-3-mediated cargo transport to dendrites. EMBO J. e201592929.

Liu, D., and Liman, E.R. (2003). Intracellular Ca2+ and the phospholipid PIP2 regulate the taste transduction ion channel TRPM5. Proc. Natl. Acad. Sci. *100*, 15160–15165.

Logan, C.Y., and Nusse, R. (2004). The Wnt signalling pathway in development and disease. Annu. Rev. Cell Dev. Biol. *20*, 781–810.

Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-Fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer *3*, 330–338.

Lowry, W.E., Blanpain, C., Nowak, J.A., Guasch, G., Lewis, L., and Fuchs, E. (2005). Defining the impact of β-catenin/Tcf transactivation on epithelial stem cells. Genes Dev. *19*, 1596–1611.

Luis, T.C., Naber, B.A.E., Roozen, P.P.C., Brugman, M.H., De Haas, E.F.E., Ghazvini, M., Fibbe,

W.E., Van Dongen, J.J.M., Fodde, R., and Staal, F.J.T. (2011). Canonical wnt signaling regulates hematopoiesis in a dosage-dependent fashion. Cell Stem Cell.

Luo, W., Cao, Y., Liao, C., and Gao, F. (2012). Diabetes mellitus and the incidence and mortality of colorectal cancer: A meta-analysis of 24 cohort studies. Color. Dis. *14*, 1307–1312.

Lutgens, M.W.M.D., Vleggaar, F.P., Schipper, M.E.I., Stokkers, P.C.F., van der Woude, C.J., Hommes, D.W., de Jong, D.J., Dijkstra, G., van Bodegraven, A.A., Oldenburg, B., et al. (2008). High frequency of early colorectal cancer in inflammatory bowel disease. Gut *57*, 1246–1251.

Lutgens, M.W.M.D., van Oijen, M.G.H., van der Heijden, G.J.M.G., Vleggaar, F.P., Siersema, P.D., and Oldenburg, B. (2013). Declining risk of colorectal cancer in inflammatory bowel disease: an updated meta-analysis of population-based cohort studies. Inflamm. Bowel Dis. *19*, 789–799.

Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F., and Boland, C.R. (2009). Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet. *76*, 1–18.

MacDonald, B.T., Tamai, K., and He, X. (2009). Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. Dev. Cell *17*, 9–26.

Mariadason, J.M., Bordonaro, M., Aslam, F., Shi, L., Kuraguchi, M., Velcich, A., and Augenlicht, L.H. (2001). Down-Regulation of β-Catenin TCF Signaling Is Linked to Colonic Epithelial Cell Differentiation. Cancer Res. *61*, 3465–3471.

El Marjou, F., Janssen, K.P., Chang, B.H.J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186–193.

Martinez-Useros, J., and Garcia-Foncillas, J. (2016). Obesity and colorectal cancer: Molecular features of adipose tissue. J. Transl. Med. *14*.

Matsuda, T., Kawanishi, M., Yagi, T., Matsui, S., and Takebe, H. (1998). Specific tandem GG to TT base substitutions induced by acetaldehyde are due to intra-strand crosslinks between adjacent guanine bases. Nucleic Acids Res. *26*, 1769–1774.

McTiernan, A. (2008). Mechanisms linking physical activity with cancer. Nat. Rev. Cancer *8*, 205–211.

Michishita, E., Nakabayashi, K., Suzuki, T., Kaul, S.C., Ogino, H., Fujii, M., Mitsui, Y., and Ayusawa, D. (1999). 5-Bromodeoxyuridine induces senescence-like phenomena in mammalian cells regardless of cell type or species. J. Biochem. *126*, 1052–1059.

Middeldorp, A., Jagmohan-Changur, S., Van Eijk, R., Tops, C., Devilee, P., Vasen, H.F.A., Hes, F.J., Houlston, R., Tomlinson, I., Houwing-Duistermaat, J.J., et al. (2009). Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol. Biomarkers Prev. *18*, 3062–3067.

Middelhoff, M., Westphalen, C.B., Hayakawa, Y., Yan, K.S., Gershon, M.D., Wang, T.C., and Quante, M. (2017). Dclk1-expressing tuft cells: Critical modulators of the intestinal niche? Am J Physiol Gastrointest Liver Physiol *313*, G285–G299.

Mojarad, E.N., Kuppen, P.J.K., Aghdaei, H.A., and Zali, M.R. (2013). The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterol. Hepatol. from Bed to Bench *6*, 120–128.

Molenaar, M., Van De Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destrée, O., and Clevers, H. (1996). XTcf-3 transcription factor mediates β-catenin-induced axis formation in xenopus embryos. Cell *86*, 391–399.

Von Moltke, J., Ji, M., Liang, H.E., and Locksley, R.M. (2016). Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit. Nature *529*, 221–225.

Morin, P.J. (2005). Claudin proteins in human cancer: Promising new targets for diagnosis and therapy. Cancer Res. *65*, 9603–9606.

Morin, P.J., Sparks, a B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science *275*, 1787–1790.

Mork, M.E., and Vilar, E. (2016). Intestinal Polyposis Syndromes. In Intestinal Polyposis Syndromes, L.A. Boardman, ed. (Cham: Springer), pp. 25–32.

Mosinger, B., Redding, K.M., Parker, M.R., Yevshayeva, V., Yee, K.K., Dyomina, K., Li, Y., and Margolskee, R.F. (2013). Genetic loss or pharmacological blockade of testes-expressed taste genes causes male sterility. Proc. Natl. Acad. Sci. *110*, 12319–12324.

Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. (1995). Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. Proc. Natl. Acad. Sci. *92*, 3046–3050.

Munkholm, P. (2003). Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment. Pharmacol. Ther. *18*, 1–5.

Nagatake, T., Fujita, H., Minato, N., and Hamazaki, Y. (2014). Enteroendocrine cells are specifically marked by cell surface expression of claudin-4 in mouse small intestine. PLoS One *9*, e90638.

Naishiro, Y., Yamada, T., Takaoka, A.S., Hayashi, R., Hasegawa, F., Imai, K., and Hirohashi, S. (2001). Restoration of Epithelial Cell Polarity in a Colorectal Cancer Cell Line by Suppression of β-catenin/T-Cell Factor 4-mediated Gene Transactivation. Cancer Res. *61*, 2751–2758.

Nakanishi, Y., Seno, H., Fukuoka, A., Ueo, T., Yamaga, Y., Maruno, T., Nakanishi, N., Kanda, K., Komekado, H., Kawada, M., et al. (2013). Dclk1 distinguishes between tumor and normal stem cells in the intestine. Nat. Genet. *45*, 98–103.

Niederreiter, L., Fritz, T.M.J., Adolph, T.E., Krismer, A.-M., Offner, F.A., Tschurtschenthaler, M., Flak, M.B., Hosomi, S., Tomczak, M.F., Kaneider, N.C., et al. (2013). ER stress transcription factor Xbp1 suppresses intestinal tumorigenesis and directs intestinal stem cells. J. Exp. Med. *210*, 2041–2056.

Nieuwenhuis, M.H., and Vasen, H.F.A. (2007). Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. Crit. Rev. Oncol. Hematol. *61*, 153–161.

Ning, Y., Wang, L., and Giovannucci, E.L. (2010). A quantitative analysis of body mass index and colorectal cancer: Findings from 56 observational studies. Obes. Rev. *11*, 19–30.

Van Noort, M., Meeldijk, J., Van Der Zee, R., Destree, O., and Clevers, H. (2002). Wnt signaling controls the phosphorylation status of β-catenin. J. Biol. Chem. *277*, 17901–17905.

Norat, T., Lukanova, A., Ferrari, P., and Riboli, E. (2002). Meat consumption and colorectal cancer risk: Dose-response meta-analysis of epidemiological studies. Int. J. Cancer *98*, 241–256.

Nowakowski, R.S., Lewin, S.B., and Miller, M.W. (1989). Bromodeoxyuridine immunohistochemical determination of the lengths of the cell cycle and the DNA-synthetic phase for an anatomically defined population. J. Neurocytol. *18*, 311–318.

Nucci, M.R., Robinson, C.R., Longo, P., Campbell, P., and Hamilton, S.R. (1997). Phenotypic and genotypic characteristics of aberrant crypt foci in human colorectal mucosa. Hum. Pathol. *28*, 1396–1407.

Nusse, R. (2018). The Wnt Homepage, http://web.stanford.edu/group/nusselab/cgi-bin/wnt. Accessed August 2018.

Nyström, E.E.L., Birchenough, G.M.H., Post, S. Van Der, Arike, L., Gruber, A.D., Hansson, G.C., and Johansson, M.E. V (2018). Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. EBioMedicine *1*, 1–10.

Oakley, B.R., Paolillo, V., and Zheng, Y. (2015). γ-Tubulin complexes in microtubule nucleation and beyond. Mol. Biol. Cell *26*, 2957–2962.

Obe, G., Jonas, R., and Schmidt, S. (1986). Metabolism of ethanol in vitro produces a compound which induces sister-chromatid exchanges in human peripheral lymphocytes in vitro: Acetaldehyde not ethanol is mutagenic. Mutat. Res. Lett. *174*, 47–51.

Ogaki, S., Shiraki, N., Kume, K., and Kume, S. (2013). Wnt and Notch signals guide embryonic stem cell differentiation into the intestinal lineages. Stem Cells *31*, 1086–1096.

Ooi, L.Y. (2016). Post-GWAS functional characterisation of colorectal cancer risk loci. PhD. University of Edinburgh.

Park, S.-W., Zhen, G., Verhaeghe, C., Nakagami, Y., Nguyenvu, L.T., Barczak, A.J., Killeen, N., and Erle, D.J. (2009). The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc. Natl. Acad. Sci. *106*, 6950–6955.

Peltekova, V.D., Lemire, M., Qazi, A.M., Zaidi, S.H.E., Trinh, Q.M., Bielecki, R., Rogers, M., Hodgson, L., Wang, M., D'Souza, D.J.A., et al. (2014). Identification of genes expressed by immune cells of the colon that are regulated by colorectal cancer-associated variants. Int. J. Cancer *134*, 2330–2341.

Peltomaki, P., Vasen, H., and Jass, J.R. (2000). Hereditary nonpolyposis colorectal cancer. In World Health Organisation Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System, S.R. Hamilton, and L.A. Aaltonen, eds. (Lyon: IARC Press), pp. 126–129.

Peltomäki, P. (2005). Lynch syndrome genes. Fam. Cancer *4*, 227–232.

Pernick, N. (2005). Stains - Mucins, http://www.pathologyoutlines.com/topic/stainsmucins.html. Accessed July 2018.

Peters, U., Bien, S., and Zubair, N. (2015). Genetic architecture of colorectal cancer. Gut *64*, 1623–1636.

Phipps, A.I., Newcomb, P.A., Garcia-Albeniz, X., Hutter, C.M., White, E., Fuchs, C.S., Hazra, A., Ogino, S., Nan, H., Ma, J., et al. (2012). Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. Gastroenterology *143*, 51–54.

Pierre, F., Freeman, A., Tache, S., Van der Meer, R., and Corpet, D.E. (2004). Beef meat and blood sausage promote the formation of azoxymethane-induced mucin-depleted foci and aberrant crypt foci in rat colons. J. Nutr. *134*, 2711–2716.

Pino, M.S., and Chung, D.C. (2014). The Chromosomal Instability Pathway in Colon Cancer. Gastroenterology *138*, 2059–2072.

Pinto, D., Robine, S., Jaisser, F., El Marjou, F., and Louvard, D. (1999). Regulatory sequences of the mouse villin gene that efficiently drive transgenic expression in immature and differentiated epithelial cells of small and large intestines. J. Biol. Chem. *274*, 6476–6482.

Pinto, D., Gregorieff, A., Begthel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev. *17*, 1709–1713.

van der Ploeg, H.P., and Hillsdon, M. (2017). Is sedentary behaviour just physical inactivity by another name? Int. J. Behav. Nutr. Phys. Act. *14*, 142.

Poulsen, M., and Bisgaard, M. (2008). MUTYH Associated Polyposis (MAP). Curr. Genomics *9*, 420–435.

Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B., and Kinzler, K.W. (1992). APC mutations occur early during colorectal tumorigenesis. Nature *359*, 235–237.

Rachmilewitz, D., Stamler, J.S., Bachwich, D., Karmeli, F., Ackerman, Z., and Podolsky, D.K. (1995). Enhanced colonic nitric oxide generation and nitric oxide synthase activity in ulcerative colitis and Crohn â€™ s disease. Gut *36*, 718–723.

Ramaekers, F., Van Niekerk, C., Poels, L., Schaafsma, E., Huijsmans, A., Robben, H., Schaart, G., and Vooijs, P. (1990). Use of Monoclonal Antibodies to Keratin 7 in the Differential Diagnosis of Adenocarcinomas. Am. J. Pathol. *136*, 641–655.

Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., and Weissman, I.L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature *423*, 409–414.

Rijcken, F.E.M., Hollema, H., and Kleibeuker, J.H. (2002). Proximal adenomas in hereditary non-polyposis colorectal cancer are prone to rapid malignant transformation. Gut *50*, 382–386.

Robsahm, T.E., Aagnes, B., Hjartå Ker, A., Langseth, H., Bray, F.I., and Larsen, I.K. (2013). Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. Eur. J. Cancer Prev. *22*, 492–505.

Saleque, S., Cameron, S., and Orkin, S.H. (2002). The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. Genes Dev. *16*, 301–306.

Sansom, O.J., Reed, K.R., Hayes, A.J., Ireland, H., Brinkmann, H., Newton, I.P., Batlle, E., Simon-Assmann, P., Clevers, H., Nathke, I.S., et al. (2004). Loss of Apc in vivo immediately

perturbs Wnt signaling, differentiation, and migration. Genes Dev. *18*, 1385–1390.

Sato, T., Vries, R.G., Snippert, H.J., Van De Wetering, M., Barker, N., Stange, D.E., Van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Sato, T., Van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., Van Den Born, M., Barker, N., Shroyer, N.F., Van De Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415–418.

Schatoff, E.M., Leach, B.I., and Dow, L.E. (2017). WNT Signaling and Colorectal Cancer. Curr. Colorectal Cancer Rep. *13*, 101–110.

Schepers, A.G., Snippert, H.J., Stange, D.E., van den Born, M., van Es, J.H., van de Wetering, M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mous intestinal adenomas. Science (80-.). *337*, 730–735.

Schoen, R.E., Weissfeld, J.L., Kuller, L.H., Thaete, F.L., Evans, R.W., Hayes, R.B., and Rosen, C.J. (2005). Insulin-like growth factor-I and insulin are associated with the presence and advancement of adenomatous polyps. Gastroenterology *129*, 464–475.

Schuijers, J., Junker, J.P., Mokry, M., Hatzis, P., Koo, B.K., Sasselli, V., Van Der Flier, L.G., Cuppen, E., Van Oudenaarden, A., and Clevers, H. (2015). Ascl2 acts as an R-spondin/wnt-responsive switch to control stemness in intestinal crypts. Cell Stem Cell *16*, 158–170.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685–700.

Shi, G.-X., Harrison, K., Wilson, G.L., Moratz, C., and Kehrl, J.H. (2002). RGS13 regulates germinal center B lymphocytes responsiveness to CXC chemokine ligand (CXCL)12 and CXCL13. J. Immunol. *169*, 2507–2515.

Shkoda, A., Ruiz, P.A., Daniel, H., Kim, S.C., Rogler, G., Sartor, R.B., and Haller, D. (2007). Interleukin-10 Blocked Endoplasmic Reticulum Stress in Intestinal Epithelial Cells: Impact on Chronic Inflammation. Gastroenterology *132*, 190–207.

Sieber, O.M., Lipton, L., Crabtree, M., Heinimann, K., Fidalgo, P., Phillips, R.K.S., Bisgaard, M.-L., Orntoft, T.F., Aaltonen, L.A., Hodgson, S. V., et al. (2003). Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations in *MYH*. N. Engl. J. Med. *348*, 791–799.

Van der Sluis, M., De Koning, B.A.E., De Bruijn, A.C.J.M., Velcich, A., Meijerink, J.P.P., Van Goudoever, J.B., Büller, H.A., Dekker, J., Van Seuningen, I., Renes, I.B., et al. (2006). Muc2-Deficient Mice Spontaneously Develop Colitis, Indicating That MUC2 Is Critical for Colonic Protection. Gastroenterology *131*, 117–129.

Smerdel, A., Dai, K.Z., Lorentzen, A.R., Flatø, B., Maslinski, S., Thorsby, E., Førre, and Spurkland, A. (2004). Genetic assocaition between juvenile rheumatoid arthritis and polymorphism in the SH2D2A gene. Genes Immun. *5*, 310–312.

Smillie, C. (2015). Functional Characterisation of the 11q23.1 Colorectal Cancer Risk Locus. PhD. University of Edinburgh.

Smith, C.G., Fisher, D., Harris, R., Maughan, T.S., Phipps, A.I., Richman, S., Seymour, M., Tomlinson, I., Rosmarin, D., Kerr, D., et al. (2015). Analyses of 7,635 patients with colorectal

cancer using independent training and validation cohorts show that rs9929218 in CDH1 is a prognostic marker of survival. Clin. Cancer Res. *21*, 3453–3461.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144.

Song, Y., Liu, M., Yang, F.G., Cui, L.H., Li, X.Y., and Chen, C. (2015). Dietary Fibre and the Risk of Colorectal Cancer: a Case- Control Study. Asian Pacific J. Cancer Prev. *16*, 3747–3752.

Special Diets Services Rat and Mouse No.3 Breeding Diet Data Sheet, http://www.sdsdiets.com/pdfs/RM3-P.pdf.

Specian, R.D., and Oliver, M.G. (1991). Functional biology of intestinal goblet cells. Am J Physiol Cell Physiol. *260*, C183–C193.

Su, L.-K., Kinzler, K.W., Vogelstein, B., Preisinger, A.C., Rapaich Moser, A., Luongo, C., Gould, K.A., and Dove, W.F. (1992). Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science (80-.). *256*, 668–670.

Talbot, I.C., Burt, R., Jarvinen, H., and Thomas, G. (2000). Familial adenomatous polyposis. In World Health Organisation Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System2, S.R. Hamilton, and L.A. Aaltonen, eds. (Lyon: IARC Press), pp. 120–125.

Talseth-Palmer, B.A., Brenne, I.S., Ashton, K.A., Evans, T.J., McPhillips, M., Groombridge, C., Suchy, J., Kurzawski, G., Spigelman, A., Lubinski, J., et al. (2010). Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in lynch syndrome. J. Med. Genet. *48*, 279–284.

Tanikawa, C., Kamatani, Y., Takahashi, A., Momozawa, Y., Leveque, K., Nagayama, S., Mimori, K., Mori, M., Ishii, H., Inazawa, J., et al. (2018). GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. Carcinogenesis *39*, 652–660.

Tanskanen, T., van den Berg, L., Välimäki, N., Aavikko, M., Ness-Jensen, E., Hveem, K., Wettergren, Y., Bexe Lindskog, E., Tõnisson, N., Metspalu, A., et al. (2018). Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. Int. J. Cancer *142*, 540–546.

Tenesa, A., Farrington, S.M., Prendergast, J.G.D., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N., et al. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet. *40*, 631–637.

Terasaki, H., Saitoh, T., Shiokawa, K., and Katoh, M. (2002). Frizzled-10, up-regulated in primary colorectal cancer, is a positive regulator of the WNT - β-catenin - TCF signaling pathway. Int. J. Mol. Med. *9*, 107–112.

Theodoratou, E., Farrington, S.M., Timofeeva, M., Din, F.V.N., Svinti, V., Tenesa, A., Liu, T., Lindblom, A., Gallinger, S., Campbell, H., et al. (2018). Genome-wide scan of the effect of common nsSNPs on colorectal cancer survival outcome. Br. J. Cancer 1–6.

Thumherr, N., Deschner, E.E., Stonehill, E.H., and Lipkin, M. (1973). Induction of Adenocarcinomas of the Colon in Mice by Weekly Injections of 1, 2-Dimethylhydrazine.

Cancer Res. *33*, 940–945.

Traber, P.G. (1999). Development of Brushborder Enzyme Activity. In Development of the Gastrointestinal Tract, I.R. Sanderson, and W.A. Walker, eds. (PMPH), pp. 103–111.

Tran, H.T., Sekkali, B., Van Imschoot, G., Janssens, S., and Vleminckx, K. (2010). Wnt/β-catenin signaling is involved in the induction and maintenance of primitive hematopoiesis in the vertebrate embryo. Proc. Natl. Acad. Sci. *107*, 16160–16165.

Treuting, P.M., and Dintzis, S.M. (2012). Lower Gastrointestinal Tract. In Comparative Anatomy and Histology: A Mouse and Human Atlas, P.M. Treuting, and S.M. Dintzis, eds. (Elsevier), pp. 177–201.

Treuting, P.M., Valasek, M.A., and Dintzis, S.M. (2012). Upper Gastrointestinal Tract. In Comparative Anatomy and Histology: A Mouse and Human Atlas, P.M. Treuting, and S.M. Dintzis, eds. (Elsevier), pp. 166–175.

Tsoi, K.K.F., Pau, C.Y.Y., Wu, W.K.K., Chan, F.K.L., Griffiths, S., and Sung, J.J.Y. (2009). Cigarette Smoking and the Risk of Colorectal Cancer: A Meta-analysis of Prospective Cohort Studies. Clin. Gastroenterol. Hepatol. *7*, 682–688.

Ueno, K., Hiura, M., Suehiro, Y., Hazama, S., Hirata, H., Oka, M., Imai, K., Dahiya, R., and Hinoda, Y. (2008). Frizzled-7 as a potential therapeutic target in colorectal cancer. Neoplasia *10*, 697–705.

Vassen, L., Khandanpour, C., Ebeling, P., Van Der Reijden, B.A., Jansen, J.H., Mahlmann, S., Dührsen, U., and Möröy, T. (2009). Growth factor independent 1b (Gfi1b) and a new splice variant of Gfi1b are highly expressed in patients with acute and chronic leukemia. Int. J. Hematol. *89*, 422–430.

Velcich, A., Yang, W., Heyer, J., Fragale, A., Nicholas, C., Viani, S., Kucherlapati, R., Lipkin, M., Yang, K., and Augenlicht, L. (2002). Colorectal cancer in mice genetically deficient in the mucin Muc2. Science (80-.). *295*, 1726–1729.

Vincan, E., Darcy, P.K., Farrelly, C.A., Faux, M.C., Brabletz, T., and Ramsay, R.G. (2007). Frizzled-7 dictates three-dimensional organization of colorectal cancer cell carcinoids. Oncogene *26*, 2340–2352.

De Visser, K.E., Eichten, A., and Coussens, L.M. (2006). Paradoxical roles of the immune system during cancer development. Nat. Rev. Cancer *6*, 24–37.

Vlad-Fiegen, A., Langerak, A., Eberth, S., and Müller, O. (2012). The Wnt pathway destabilizes adherens junctions and promotes cell migration via β-catenin and its target gene cyclin D1. FEBS Open Bio *2*, 26–31.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Smits, A.M.M., and Bos, J.L. (1988). Genetic Alterations during Colorectal-Tumor Development. N. Engl. J. Med. *319*, 525–532.

Voloshanenko, O., Erdmann, G., Dubash, T.D., Augustin, I., Metzig, M., Moffa, G., Hundsrucker, C., Kerr, G., Sandmann, T., Anchang, B., et al. (2013). Wnt secretion is required to maintain high levels of Wnt activity in colon cancer cells. Nat. Commun. *4*, e2610.

Walko, C.M., and Lindley, C. (2005). Capecitabine: a review. Clin. Ther. *27*, 23–44.

Weiss, A., and Attisano, L. (2013). The TGFbeta superfamily signaling pathway. Wiley Interdiscip. Rev. Dev. Biol. *2*, 47–63.

Wijnen, J.T., Brohet, R.M., van Eijk, R., Jagmohan–Changur, S., Middeldorp, A., Tops, C.M., van Puijenbroek, M., Ausems, M.G.E.M., Gómez García, E., Hes, F.J., et al. (2009). Chromosome 8q23.3 and 11q23.1 Variants Modify Colorectal Cancer Risk in Lynch Syndrome. Gastroenterology *136*, 131–137.

Win, A.K., Hopper, J.L., Buchanan, D.D., Young, J.P., Tenesa, A., Dowty, J.G., Giles, G.G., Goldblatt, J., Winship, I., Boussioutas, A., et al. (2013). Are the common genetic variants associated with colorectal cancer risk for DNA mismatch repair gene mutation carriers? Eur. J. Cancer *49*, 1578–1587.

Xiong, F., Wu, C., Bi, X., Yu, D., Huang, L., Xu, J., Zhang, T., Zhai, K., Chang, J., Tan, W., et al. (2010). Risk of Genome-Wide Association Study-Identified Genetic Variants for Colorectal Cancer in a Chinese Population. Cancer Epidemiol. Biomarkers Prev. *19*, 1855–1861.

Xiong, L., Wen, Y., Miao, X., and Yang, Z. (2014). NT5E and FcGBP as key regulators of TGF-1-induced epithelial-mesenchymal transition (EMT) are associated with tumor progression and survival of patients with gallbladder cancer. Cell Tissue Res. *355*, 365–374.

Xue, K., Li, F.-F., Chen, Y.-W., Zhou, Y.-H., and He, J. (2017). Body mass index and the risk of cancer in women compared with men. Eur. J. Cancer Prev. *26*, 94–105.

Yang, B., Cao, L., Liu, B., McCaig, C.D., and Pu, J. (2013). The Transition from Proliferation to Differentiation in Colorectal Cancer Is Regulated by the Calcium Activated Chloride Channel A1. PLoS One *8*, e60861.

You, S., Ohmori, M., Peña, M.M.O., Nassri, B., Quiton, J., Al-Assad, Z.A., Liu, L., Wood, P.A., Berger, S.H., Liu, Z., et al. (2006). Developmental abnormalities in multiple proliferative tissues of Apc Min/+ mice. Int. J. Exp. Pathol. *87*, 227–236.

Young, B., Lowe, J.S., Stevens, A., and Heath, J. (2006). Wheater's Functional Histology: A Text and Colour Atlas (Edinburgh: Churchill Livingstone Elsevier).

Zhang, H.-S., Chen, Y., Fan, L., Xi, Q.-L., Wu, G.-H., Li, X.-X., Yuan, T.-L., He, S.-Q., Yu, Y., Shao, M.-L., et al. (2015). The Endoplasmic Reticulum Stress Sensor IRE1α in Intestinal Epithelial Cells Is Essential for Protecting against Colitis. J. Biol. Chem. *290*, 15327–15336.

Zhang, W., Liu, H.T., and Tu LIU, H. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res. *12*, 9–18.

Zhao, F., Edwards, R., Dizon, D., Afrasiabi, K., Mastroianni, J.R., Geyfman, M., Ouellette, A.J., Andersen, B., and Lipkin, S.M. (2010). Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in Agr2-/- mice. Dev. Biol. *338*, 270–279.

Zheng, W., Rosenstiel, P., Huse, K., Sina, C., Valentonyte, R., Mah, N., Zeitlmann, L., Grosse, J., Ruf, N., Nürnberg, P., et al. (2006). Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. Genes Immun. *7*, 11–18.

Zoltewicz, J.S., Ashique, A.M., Choe, Y., Lee, G., Taylor, S., Phamluong, K., Solloway, M., and Peterson, A.S. (2009). Wnt signaling is regulated by endoplasmic reticulum retention. PLoS One *4*, e6191.